

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: December 1, 2005, 08:15:30 ; Search time 189 Seconds  
(without alignments)  
378.935 Million cell updates/sec

Title: US-09-924-340-58

Sequence: 1 MGPPEFGKGTGHGDLGPKG.....GAMPMEQGYPMKTMKGPFG 163

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

## Database :

A.GeneSeq\_21:.\*  
1: geneseqp1980s:.\*  
2: geneseqp1990s:.\*  
3: geneseqp2000s:.\*  
4: geneseqp2001s:.\*  
5: geneseqp2002s:.\*  
6: geneseqp2003as:.\*  
7: geneseqp2003bs:.\*  
8: geneseqp2004s:.\*  
9: geneseqp2005s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	952	100.0	163	ABR48481	Abt48481 Human Alp
2	840.5	88.3	1603	ADQ21405	Adq21405 Human Alp
3	620	65.1	218	ADL25758	Adl25758 Human can
4	419	44.0	330	AAW57645	AAW57645 Collagen-
5	419	44.0	408	AAW07539	AAW07539 Collagen-
6	419	44.0	408	AB64007	AB64007 CLP prote
7	419	44.0	408	AAW2737	AAW2737 Repetitiv
8	412.5	43.3	357	AAW5115	AAW5115 Interveni
9	412.5	43.3	357	AAW57646	AAW57646 Collagen-
10	412.5	43.3	357	AAW64008	AAW64008 CLP-CB pr
11	412.5	43.3	357	AAW72738	AAW72738 Repetitiv
12	399.5	42.0	1071	ADP94310	ADP94310 Human col
13	395	41.5	67	ABR01788	ABR01788 Human bre
14	393.5	41.3	503	ADW5808	ADW5808 Chicken t
15	390.5	41.0	553	ADP69307	ADP69307 Human lun
16	390	41.0	546	AAW40115	AAW40115 Human alp
17	390	41.0	546	ABP43633	ABP43633 IV collagen
18	390	41.0	910	ADQ19402	ADQ19402 Human sof
19	390	41.0	1268	ABO84442	ABO84442 Human can
20	390	41.0	1283	ABO84443	ABO84443 Human can
21	390	41.0	1685	ABG04839	ABG04839 Novel hum
22	390	41.0	1685	ABO84444	ABO84444 Human can
23	390	41.0	1693	ABG15619	ABG15619 Novel hum
24	389.5	40.9	1064	AAW93254	AAW93254 Collagen-

25	389.5	40.9	1064	2	AAW57652	AAW57652 Collagen-
26	389.5	40.9	1065	2	AAW37741	AAW37741 Collagen-
27	388.5	40.8	561	2	AAW37739	AAW37739 Collagen-
28	388.5	40.8	561	2	AAW93249	AAW93249 Collagen-
29	388.5	40.8	561	2	AAW57650	AAW57650 Collagen-
30	388	40.8	1497	2	AAW43711	AAW43711 Bullous p
31	386.5	40.6	252	8	ABO84441	ABO84441 Mouse can
32	386	40.5	252	2	AAW37743	AAW37743 Collagen-
33	385	40.4	252	2	AAW37738	AAW37738 Collagen-
34	385	40.4	252	2	AAW93248	AAW93248 Collagen-
35	385	40.4	1497	6	ADW83846	ADW83846 Human COL
36	384	40.3	1466	4	AAE02534	AAE02534 Bovine al
37	384	40.3	1466	4	AAE02533	AAE02533 Bovine al
38	383.5	40.3	252	2	AAW01418	AAW01418 Collagen-
39	383.5	40.3	1838	2	AAW53257	AAW53257 Human col
40	383.5	40.3	1838	7	ADW55566	ADW55566 Human pro
41	383.5	40.3	1838	7	ADW55570	ADW55570 Human pro
42	383.5	40.3	1838	7	ADP65220	ADP65220 Human alp
43	383.5	40.3	1838	9	ADW70238	ADW70238 Tumor-aa
44	383.5	40.3	1838	9	ADZ09822	ADZ09822 Human bre
45	383.5	40.3	1838	9	ADZ70492	ADZ70492 Human pro
46	382	40.1	638	8	ADW08295	ADW08295 Human NOV
47	381.5	40.1	703	6	ABP96315	ABP96315 Human COL
48	381.5	40.1	703	8	ADW98189	ADW98189 Protein f
49	381.5	40.1	717	6	ABP96314	ABP96314 Human COL
50	381.5	40.1	733	6	ABU69145	ABU69145 Human NOV
51	381.5	40.1	733	8	ADW08293	ADW08293 Human NOV
52	381.5	40.1	1181	8	ADP22968	ADP22968 PRO polyp
53	380.5	40.0	638	8	ABU69146	ABU69146 Human NOV
54	380	39.9	1767	6	ADQ39813	ADQ39813 Human myo
55	380	39.9	1767	8	ADQ39817	ADQ39817 Human myo
56	380	39.9	1806	8	ADQ39816	ADQ39816 Human myo
57	380	39.9	1806	8	ADQ39815	ADQ39815 Human myo
58	380	39.9	1818	8	ADQ39812	ADQ39812 Human myo
59	380	39.9	1818	8	ADQ39814	ADQ39814 Human myo
60	378.5	39.8	635	4	AAW78798	AAW78798 Human pro
61	378.5	39.8	635	4	AAW78798	AAW78798 Human pro
62	378.5	39.8	644	4	AAW79782	AAW79782 Human pro
63	378.5	39.8	644	4	ADW98659	ADW98659 Protein f
64	378.5	39.8	1745	4	AAW7793	AAW7793 Human pro
65	378.5	39.8	1745	4	ABW97234	ABW97234 Novel hum
66	378.5	39.8	1745	8	ADQ19841	ADQ19841 Human sof
67	377.5	39.7	1466	8	AAE02537	AAE02537 Porcine a
68	377	39.6	1028	8	ADN35278	ADN35278 Helical d
69	377	39.6	1078	2	AAW71704	AAW71704 Collagen
70	377	39.6	1078	3	AAW96125	AAW96125 Collagen
71	377	39.6	1078	5	AAE16478	AAE16478 Human COL
72	377	39.6	1078	5	ABW80736	ABW80736 Collagen
73	377	39.6	1078	5	ABW809628	ABW809628 Amino aci
74	377	39.6	1078	7	ADP13078	ADP13078 Human COL
75	377	39.6	1313	8	ADN35279	ADN35279 Synthetic
76	377	39.6	1313	8	ADN35277	ADN35277 Helical d
77	377	39.6	1449	4	AAE02535	AAE02535 Porcine a
78	377	39.6	1466	4	ABW50291	ABW50291 Collagen
79	377	39.6	1466	5	ABW90747	ABW90747 Human tun
80	377	39.6	1466	5	ABW54454	ABW54454 Human tun
81	377	39.6	1466	6	ABW47418	ABW47418 Breast ca
82	377	39.6	1466	6	ADP65248	ADP65248 Human alp
83	377	39.6	1466	7	ADP65210	ADP65210 Human alp
84	377	39.6	1466	8	ADQ26091	ADQ26091 Type III,
85	377	39.6	1466	8	ADQ26677	ADQ26677 Human COL
86	377	39.6	1466	8	ADP16802	ADP16802 Human COL
87	377	39.6	1466	8	ADP16427	ADP16427 Human COL
88	377	39.6	1466	8	ABW80366	ABW80366 Tumor-aa
89	377	39.6	1466	8	ADP67267	ADP67267 Human bla
90	377	39.6	1466	9	ADW70235	ADW70235 Tumor-aa
91	377	39.6	1466	9	ADZ09873	ADZ09873 Human bre
92	377	39.6	1466	9	AAE04495	AAE04495 Human pro
93	377	39.6	1469	4	ABG15191	ABG15191 Novel hum
94	377	39.6	1470	7	ADW09339	ADW09339 Novel pro
95	377	39.6	1572	8	ADU04510	ADU04510 SPTI-COLL
96	377	39.6	1604	8	ADU04497	ADU04497 Modified
97	377	39.6	1726	6	ABW42661	ABW42661 Decorin-m

	377	39.6	1950	8	ADU04493	Adu04493 Modified
98	377	39.6	1950	8	ADU04493	Adu04493 Modified
99	375.5	39.4	231	2	AAy23937	Aay23937 Amino aci
100	374	39.3	1518	4	ABG22679	Abg22679 Novel hum

## ALIGNMENTS

Query Match	100.0%	Score 952;	DB 6;	Length 163;
Best Local Similarity	100.0%	Pred. No. 1.3e-63;		
Matches 163;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
1	MGPFGFKKTGHPGLPGPKGDCGKGPFGSTGRPGAEGPGAMGPQGRPPGAVGPPGP	60		
1	MGPFGFKKTGHPGLPGPKGDCGKGPFGSTGRPGAEGPGAMGPQGRPPGAVGPPGP	60		
61	PGQPGPAGISAVGLKGRGATGERGLAGLPGQPGPPGPGPGPGYKMGATGPMGQGIIPG	120		
61	PGQPGPAGISAVGLKGRGATGERGLAGLPGQPGPPGPGPGPGYKMGATGPMGQGIIPG	120		
121	IPGPPGPMGPGFGKAGHCNPSDCFGAMPMBQOYIPPMKTKMGPPG	163		
121	IPGPPGPMGPGFGKAGHCNPSDCFGAMPMBQOYIPPMKTKMGPPG	163		

RESULT 2  
ADQ21405  
ID ADQ21405 standard; protein; 1603 AA

DT	26-AUG-2004	(first entry)
DE	Human soft tissue sarcoma-upregulated protein - SEQ ID 4225.	
KW	soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.	
XX	Homo sapiens.	
OS		
PN	WO2004048938-A2.	
PD	10-JUN-2004.	
XX		
PF	26-NOV-2003; 2003WO-US038193.	
PR	26-NOV-2002; 2002US-0429739P.	
PA	(PROT-) PROTEIN DESIGN LABS INC.	
PI	Aziz N, Ginsburg WM, Zlotnick A;	
DR	WPI; 2004-441208/41.	
XX		
PT	Early detection of soft tissue sarcoma comprises determining expression	
PT	of a gene in a first soft tissue sample and a normal soft tissue sample	
PT	and comparing the gene expression, also useful in treating soft tissue	
PS	sarcoma.	
XX		
PS	Example 2; SEQ ID NO 4225; 210pp; English.	
XX		
CC	The invention relates to a novel method for detecting soft tissue sarcoma	
CC	which comprises obtaining a first soft tissue sample from an individual,	
CC	and a normal soft tissue sample from the same or different individual,	
CC	determining the expression of a gene in both samples and comparing the	
CC	expression of the gene in both soft tissue samples, where a higher level	
CC	of protein expression in the first soft tissue sample indicates the	
CC	presence of soft tissue sarcoma. The method of the invention has	
CC	cytostatic applications and may be useful for detecting soft tissue	
CC	sarcoma, possibly via gene therapy or vaccine production. The nucleic	
CC	acid sequences may be useful in diagnostic and screening applications.	
CC	The current sequence is that of a human soft tissue sarcoma-upregulated	
CC	protein of the invention. The current sequence is not shown within the	
XX	specification per se but was submitted in CD format by the inventor.	
XX		
SQ	Sequence 1603 AA:	
Query Match	88.3%; Score 840.5; DB 8; Length 1603;	
Best Local Similarity	44.5%; Pred. No. 2.2e-54;	
Matches 163; Conservative	0; Mismatches 0; Indels 203; Gaps 1;	
QY	1 MGPFGFKKTHPPGLPGPKGDCCKRPGPGSGRGRRGASGEPEGMPOGRPGPHGVGGP 60	
DB	1238 MGPFGFKKTHPPGLPGPKGDCCKRPGPGSGRGRRGASGEPEGMPOGRPGPHGVGGP 1237	
QY	61 PGQPGAGISAVGAKGDRGATGERGLAGLPQQ----- 92	
DB	1238 PGQPGAGISAVGAKGDRGATGERGLAGLPQQPGPPGHPPRGHPPGTGDGAAGKGSPGKQ 1357	
QY	93 ----- 92	
DB	1358 GFTYGPDPKGDPGAAGQGQAGEKGRAGAMPGPSGSMPGVGPDPAGERGHGAPGPS 1417	
QY	93 ----- 92	
DB	1418 GSPGIPGVPGSMGMVVNYDEIKREIRBEIIMFDERMAVYTSMQPFEMAAADGRDPP 1477	



Dd	217	GPFGAGCVGSPGAPRPPGPPPGPCAPPBPDPGGPFPAGPVGSFAGAPCPPP	276
Oy	114	GQDGIPIGIPGPPGMGPGKAGHCNPSDCFAM	146
Dd	277	GPFGPFGARPPGPPGPPGPPGAPGVSGPAM	309
RESULT 5 AAW07539 standard; protein; 408 AA.			
XX	ID	AAW07539	
AC	AAW07539;		
XX	25-MAR-2003	(revised)	
DT	03-FEB-1997	(first entry)	
XX			
DE	Collagen like protein (CLP).		
XX			
KW	Polymer; repeat unit; natural collagen; intervening oligopeptide; fibre;		
KM	film; membrane; emulsion; coating; collagen like protein;		
RW	specific binding material; catalyst; purification agent; composite;		
KW	laminare; adhesive; cell growth surface; affinity column;		
KW	biological material support; wound dressing; in vivo prothesis.		
XX	Synthetic.		
OS			
PN	US5514581-A.		
PD	07-MAY-1996.		
PF	06-NOV-1990;	90US-00609716.	
PR	04-NOV-1986;	86US-00927258.	
PR	29-OCT-1987;	87US-00114618.	
PR	09-NOV-1988;	88US-00269429.	
PR	07-NOV-1989;	89WO-US005016.	
XX	(PROT-) PROTEIN POLYMER TECHNOLOGIES INC.		
PA	Cappello J, Ferrari FA;		
PI	WP1; 1996-238772/24.		
XX			
DR	DNA encoding protein contg. repeated fibroin derived segments - linked by		
PT	oligopeptide with cell adhesion properties useful, e.g. in wound		
PT	dressings.		
XX			
XX	Example 3; Col 99-102; 71pp; English.		
PS			
XX			
CC	Novel DNA sequence encodes a polymer comprising segments of repeating		
CC	units of 3-9 amino acids from natural collagen, i.e. the present sequence		
CC	collagen like protein (CLP), able to assemble into aligned structures		
CC	formable into articles. The polymer comprises at least 2 segments joined		
CC	by an unaligned intervening oligopeptide, other than the repeating unit.		
CC	The polymer can be used to make fibres, films, membranes, emulsions,		
CC	coatings, etc., useful as, e.g. specific binding materials, catalysts,		
CC	puffing agents, composites, laminates, adhesives, cell growth surfaces,		
CC	affinity columns and supports for biological materials. Typical		
CC	applications include wound dressings, and in vivo protheses. The polymer		
CC	produces articles with good mechanical properties, and the intervening		
CC	oligopeptide can provide a ligand for binding a mol., antibody, etc., or		
CC	a chemically reactive site for coupling to proteins, etc.. (Updated on 25-		
CC	MAR-2003 to correct PF field.)		
XX			
SO	Sequence 408 AA;		
Query Match	44.0%;	Score 419;	DB 2; Length 408;
Best Local Similarity	54.2%;	Pred. No. 1.8e-23;	
Matches	83; Conservative	2; Mismatches	60; Indels 8; Gaps 2.
Oy	2	GPFGKGTGHPLGFPGYDCGKPGPPPGSTGRPGAEGEPGAMGPGRRPGPHGVPPPP	61

D6		235	GPPGPRGPARGPGRPGRPCRPBPAPBVGSGAAGRRPGRGGRCRAPARPPPGRPP	294
Oy		62	GQPGDAG-ISAVALKGLDRGAIGERGLAGLPGQPGRGPOGPG-----YGMATGM	113
D6		295	GPSGACGVGSFRGARGPGRPPCRPPCPGARPGRRPGSPGPPGPAVBVSGAFGPPGPP	354
Oy		114	GQQGI RGI RGPRGPWGQRPKAKHCHNPSDFGAM	146
D6		355	GPDPGPAPGDPPGPDPGPDPGPAPGBVGSFGAM	387
 RESULT 6 AAB64007				
ID	AAB64007	standard; protein; 408 AA.		
XX AC	AAB64007;			
XX DT	19-MAR-2001	(first entry)		
XX XX	CLP	protein sequence SEQ ID 65.		
KW KM	Proteinaceous polymer; repeat unit; structural polymer; coating; film; fibre; membrane; adhesive; emulsion; laminate; keratin; collagen.			
OS	Synthetic.			
PX PN	US6140072-A.			
PD PD	31-OCT-2000.			
PF PF	07-JUN-1995;	95US-00475411.		
PR PR	04-NOV-1986;	86US-00927258. 29-OCT-1987;		
PR PM	09-NOV-1988;	87US-00114618. 88US-00269429.		
PR PX	06-NOV-1990;	90US-00609716.		
PA PA	(PROT-) PROTEIN POLYMER TECHNOLOGIES INC.			
PI PI	Cappello J, Ferrari FA;			
DR DR	WI:	2001-048958/06.		
PT PT	New DNA encoding a polymer with strands of repeating units of natural weight polymers joined by intervening oligopeptide for producing high molecular weight polymers of amino acids.			
PS PS	Example 2; Col 41; 73pp; English.			
XX XX	This invention relates to DNA encoding a proteinaceous polymer. The polymer comprises strands of repeating units of a natural protein capable of assembling into aligned structures, with at least 2 strands joined by an intervening oligopeptide other than the repeating units. The intervening oligopeptide is unaligned and the polymer has individual strands of the same or different repeating units. The DNA is useful for producing high molecular weight polymers of amino acids based on biologically and chemically active structural polymers. These polymers may be used to provide a variety of structures for different purposes, and to produce articles including coatings, or other (non)structural components, e.g. fibres, films, membranes, adhesives or emulsions, or with other compounds and/or compositions to form composites or laminates. Peptide sequences AAB63j71-B63j91 represent monomer sequences which can be used in the polymers of the invention. Oligonucleotide sequences CC AAPF3370 - AAPF23386 and amino acid sequences AAB63992 - AAB64002 are used in the construction of SLP and PCB-SLP polymers. Oligonucleotide sequences AAPF3387 - AAPF23397 and amino acid sequences AAB64003 - AAB64008 are used in the construction of CLP (collagen like protein) polymers. Oligonucleotide sequences AAPF23398 - AAPF3409 and amino acid sequences AAB64009 - AAB64014 are used in the construction of keratin polymers. Proteins and peptides represented by sequences AAB64015 - AAB64049 are examples of polymers of the invention			
XX XX	Sequence 408 AA;			

Query Match 44.0%; Score 419; DB 4; Length 408;  
 Best Local Similarity 54.2%; Pred. No. 1.8e-23;  
 Matches ... 83; Conservative 2; Mismatches 60; Indels 8; Gaps 2;

QY 2 GPPGFKGKTGHPGLPGRPKDCCGKPPGGSTGRPGAGEPGAMCPQGRPGPHGVSPGPP 61  
 DB 235 GPPGPPGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 294  
 QY 62 GGGPGAG-1SAVGLKDDKDRATGERGLAGLPGRGPPGPPGPPGPPGPPGPPGPP 113  
 DB 295 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 354  
 QY 114 GGGGIPGIRPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 146  
 DB 355 GPPGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 387

# RESULT 7

AAB72737  
 ID AAB72737 standard; protein; 408 AA.

AC AAB72737;  
 XX  
 DT 09-MAY-2001 (first entry)  
 DE Repetitive protein polymer protein sequence #37.

XX Protein polymer; repetitive sequence; repeat; mechanical property;  
 KW coating; fibre; film; membrane; adhesive; emulsion; composite; laminate;  
 KM diagnosis; sensor.

XX Unidentified.

OS US6184348-B1.

PN 06-FEB-2001.

PD 07-JUN-1995; 95US-00478029.

XX 04-NOV-1986; 86US-00927258.

PR 29-OCT-1987; 87US-00114618.

PR 09-NOV-1988; 88US-00269429.

PR 06-NOV-1990; 90US-00609716.

XX (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.

PI Ferrari FA, Cappello J;

XX WPI; 2001-217800/22.

DR New recombinant proteinaceous polymers comprising strands of repeating

PT units of natural protein which can be assembled into aligned structures,

PT useful for producing e.g. fibers, films, membranes, adhesives, or

PT emulsions.

XX Example 3; Col 99-102; 72pp; English.

XX The present invention provides compositions containing proteinaceous

CC polymers comprising proteins with strands of repeating units of a natural

CC protein. These can be used as coatings, fibres, films, membranes,

CC adhesives, emulsions, in composites and laminates, and in affinity

CC columns, diagnostic devices and sensors. The present sequence is a

CC protein sequence used as a demonstration of the invention

XX Sequence 408 AA;

QY Query Match 44.0%; Score 419; DB 4; Length 408;

Best Local Similarity 54.2%; Pred. No. 1.8e-23;

Matches 83; Conservative 2; Mismatches 60; Indels 8; Gaps 2;

2 GPPGFKGKTGHPGLPGRPKDCCGKPPGGSTGRPGAGEPGAMCPQGRPGPHGVSPGPP 61

|||||

DB 235 GPPGPPGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 294

QY 62 GGGPGAG-1SAVGLKDDKDRATGERGLAGLPGRGPPGPPGPPGPPGPPGPPGPP 113

DB 295 GPPGPPGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 354

QY 114 GGGGIPGIRPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 146

DB 355 GPPGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 387

# RESULT 8

AAR95115  
 ID AAR95115 standard; protein; 357 AA.

AC AAR95115;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 03-FEB-1997 (first entry)

DE Intervening sequence contg. collagen like protein (CLP)-CB.

XX Polymer; repeat unit; natural collagen; intervening oligopeptide; fibre;

KW film; membrane; emulsion; coating; collagen like protein;

KW specific binding material; catalyst; purification agent; composite;

KW laminate; adhesive; cell growth surface; affinity column;

KW biological material support; wound dressing; in vivo prosthesis.

OS Synthetic.

PN US5514581-A.

PD 07-MAY-1996.

PF 06-NOV-1990; 90US-00609716.

XX 04-NOV-1986; 86US-00927258.

PR 29-OCT-1987; 87US-00114618.

PR 09-NOV-1988; 88US-00269429.

PR 07-NOV-1989; 89WO-US005016.

XX (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.

PI Cappello J, Ferrari FA;

XX WPI; 1996-238772/24.

DR DNA encoding protein contg. repeated fibroin derived segments - linked by

XX dressings.

XX Example 3; Col 101-104; 71pp; English.

XX Novel DNA sequence encodes a polymer comprising segments of repeating

CC units of 3-9 amino acids from natural collagen, i.e. the present

CC intervening sequence contg. collagen like protein (CLP)-CB, able to

CC assemble into aligned structures formable into articles. The polymer

CC comprises at least 2 segments joined by an unaligned intervening

CC oligopeptide, other than the repeating unit. The polymer can be used to

CC make fibres, films, membranes, emulsions, coatings, etc., useful as, e.g.

CC specific binding materials, catalysts, purificn. agents, composites,

CC laminates, adhesives, cell growth surfaces, affinity columns and supports

CC for biological materials. Typical applications include wound dressings,

CC and in vivo protheses. The polymer produces articles with good mechanical

CC properties, and the intervening oligopeptide can provide a ligand for

CC binding a mol., antibody, etc., or a chemically reactive site for

CC coupling to proteins, etc.. (Updated on 25-MAR-2003 to correct PF field.)

XX Sequence 357 AA;

QY Query Match 43.3%; Score 412.5; DB 2; Length 357;

Best Local Similarity 42.7%; Pred. No. 4.9e-23;

Matches 90; Conservative 6; Mismatches 54; Indels 61; Gaps 6;





CC with function superiority to the native collagen.

XX Sequence 1071 AA;

Query Match 42.0%; Score 399.5; DB 7; Length 1071;  
Best Local Similarity 48.9%; Pred. No. 1.2e-21;  
Matches 88; Conservative 1; Mismatches 68; Indels 23; Gaps 6;

QY 2 GPPGFKGKTGHP---GLPGPKDCKPGPPGSGTGRPGAGEGFGAMGP---QGRPPGPHV 55  
DB 865 GPPGAPGAPGPPGSAGAPGPPGAPGPPGAGAPGPPGAGAPGPPGAGAPGPPGSA 924  
QY 56 GPPGPPGQPPGPA---ISAVGLKDDRGATGGERGLAGIPGCGPPGPPGPPGPPGPPG 104  
DB 925 GAPGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 984  
QY 105 GKMGGATGPMGQQGIGIPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 163  
DB 985 GAPGAPGPPGSAGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPGAPGPPG 1039

#### RESULT 13

ABR01788  
ID ABR01788 standard; protein; 67 AA.

XX ABR01788;

DT 22-APR-2003 (first entry)

XX Human breast specific polypeptide #103.

XX Human; breast specific nucleic acid; BSNM; breast; cytostatic;  
KW gene therapy; vaccines; lung cancer; breast cancer;  
KM breast specific polypeptide; BSP.

XX Homo sapiens.

XX WO200268645-A2.

XX 06-SEP-2002.

XX 20-NOV-2001; 2001WO-US045151.

XX 20-NOV-2000; 2000US-024992P.

XX (DIAD-) DIADEXUS INC.

PI Salceda S, Macina RA, Recipon H, Cafferkey R, Sun Y, Liu C;  
PI Turner LR;

DR WPI; 2002-713379/77.

PT New breast specific genes and proteins, useful in gene therapy or as  
PT vaccines for treating breast cancer or non-cancerous breast diseases, as  
PT well as for diagnosing, monitoring or staging these diseases.

XX Claim 11; Page 277; 277p; English.

CC The invention relates to a novel isolated breast specific nucleic acid  
CC molecule. The polypeptides of the invention have cytostatic activity. The  
CC as nucleic acids and polypeptides may have a use in gene therapy, and  
CC as vaccines. The breast specific nucleic acid and polypeptide are useful  
CC for diagnosing and monitoring the presence and metastases of lung cancer  
CC in a patient. The antibody that specifically binds to the breast specific  
CC polypeptide is useful for determining the presence of a breast specific  
CC protein in a sample, as well as for treating a patient with breast  
CC cancer, particularly by inducing an immune response against the breast  
CC cancer cell expressing the breast specific nucleic acid molecule or  
CC polypeptide. In particular, these breast specific genes and proteins are  
CC useful for identifying, diagnosing, monitoring, staging, imaging and  
CC treating breast cancer and non-cancerous disease states in breast tissue.  
CC These are also useful in gene therapy, production of transgenic animals  
CC and cells, and in the production of engineered breast tissue for

CC treatment and research. The sequences shown in ABR01686-ABR01788  
CC represent the novel human breast specific polypeptides of the invention

XX Sequence 67 AA;

Query Match 41.5%; Score 395; DB 5; Length 67;  
Best Local Similarity 100.0%; Pred. No. 2.2e-22;  
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 GPPGPPGKMGATGTPMGGQIGIPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 156  
DB 1 GPGGPPGKMGATGTPMGGQIGIPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 60  
QY 157 TMKGPPG 163  
DB 61 TMKGPPG 67

#### RESULT 14

ADV95808  
ID ADV95808 standard; protein; 503 AA.

XX ADV95808;

DT 10-MAR-2005 (first entry)

XX Chicken type IX collagen alpha 1 precursor.

XX arthritis; collagen type IX alpha 1 chain; antiarthritic;  
KW antiinflammatory; immunosuppressive; musculoskeletal disease;  
KM degeneration.

XX Gallus sp.

XX WO2004110475-A1.

XX 23-DEC-2004.

XX 17-JUN-2004; 2004WO-AU000788.

XX 17-JUN-2003; 2003AU-00903037.

XX (NUTR-) INST NUTRACEUTICAL RES PTY LTD.

XX Ghosh P;

DR WPI; 2005-039981/04.

PT Pharmaceutical composition for treating or preventing arthritis or other  
PT degenerative disease in an individual, comprises a polypeptide comprising  
PT a collagen type IX alpha 1 chain NC4 domain.

PS Disclosure; SEQ ID NO 17; 92p; English.

CC This invention describes a novel pharmaceutical composition for treating  
CC or preventing arthritis or other degenerative disease which comprises a  
CC polypeptide a collagen type IX alpha 1 chain NC4 domain or a biologically  
CC active fragment having antiarthritic, antiinflammatory and  
CC immunosuppressive activity in combination with a carrier. The invention  
CC describes two methods for recovering a polypeptide having anti-arthritic  
CC or anti-inflammatory activity. The first method comprises isolating a  
CC mixture comprising a GAG-peptide and a polypeptide having a molecular  
CC weight of less than 30000 Da by autolysis from connective tissue,  
CC separating the GAG-peptide from the polypeptide, and recovering the  
CC polypeptide. The second method comprises incubating connective tissue in  
CC an autolysis medium that provides a buffered pH range of 2.5-8.5 for a  
CC time and under conditions sufficient to release a GAG-peptide and a  
CC polypeptide having a molecular weight of less than 30000 Da, recovering a  
CC mixture comprising the GAG-peptide and polypeptide from the autolysis  
CC medium, separating the polypeptide from the GAG-peptide, and recovering  
CC the polypeptide having a molecular weight of less than 30000 Da. The  
CC recovered anti-arthritic or anti-inflammatory polypeptides are useful for  
CC inducing cartilage formation or for preparing a medicament for the





**THIS PAGE BLANK (USF10)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2005, 08:19:37 ; Search time 46 Seconds  
(without alignments)  
292.959 Million cell updates/sec

Title: US-09-924-340-58  
Perfect score: 952  
Sequence: 1 MGPPGFKGTGHPGLPGPKG.....GAMPMEQOYPMKTKMGPRG 163

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues  
Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/iaa/5\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/iaa/6\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/iaa/H\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/iaa/PCTUS\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/iaa/RE\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/iaa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	952	100.0	163	2	US-10-000-489-58	Sequence 58, App1
2	840.5	88.3	1603	2	US-09-949-016-6136	Sequence 6136, Ap
3	840.5	88.3	1609	2	US-09-949-016-10910	Sequence 10910, A
4	419	44.0	330	1	US-08-642-255-32	Sequence 32, App1
5	419	44.0	408	1	US-07-609-716-65	Sequence 65, App1
6	419	44.0	408	2	US-08-475-411A-65	Sequence 65, App1
7	419	44.0	408	2	US-08-478-029A-65	Sequence 65, App1
8	412.5	43.3	357	1	US-07-609-716-66	Sequence 66, App1
9	412.5	43.3	357	1	US-08-642-255-33	Sequence 66, App1
10	412.5	43.3	357	2	US-08-475-411A-66	Sequence 66, App1
11	412.5	43.3	357	2	US-08-478-029A-66	Sequence 66, App1
12	390	41.0	546	1	US-08-494-168-10	Sequence 10, App1
13	390	41.0	1268	2	US-09-949-016-7487	Sequence 7487, Ap
14	389.5	40.9	1064	1	US-08-642-255-62	Sequence 62, App1
15	388.5	40.8	504	2	US-09-219-849-3	Sequence 3, App1
16	388.5	40.8	561	1	US-08-642-255-52	Sequence 52, App1
17	386	40.5	252	1	US-08-642-255-61	Sequence 61, App1
18	385	40.4	234	1	US-08-642-255-51	Sequence 51, App1
19	378.5	39.8	638	2	US-10-001-887-108	Sequence 108, App
20	378.5	39.8	1745	2	US-09-795-061-4	Sequence 4, App1
21	378.5	39.8	1745	2	US-09-949-002-405	Sequence 405, App
22	378.5	39.8	1771	2	US-09-949-002-492	Sequence 492, App
23	377	39.6	1057	2	US-08-931-820-4	Sequence 4, App1
24	377	39.6	1078	2	US-08-963-825-21	Sequence 21, App1
25	377	39.6	1078	2	US-09-500-811-21	Sequence 21, App1
26	377	39.6	1078	2	US-09-570-573-21	Sequence 21, App1
27	377	39.6	1078	2	US-09-548-608-21	Sequence 21, App1

28	376.5	39.5	1690	2	US-09-949-016-5884	Sequence 5884, Ap
29	372.5	39.1	417	1	US-08-175-155-69	Sequence 69, App1
30	372.5	39.1	417	1	US-08-477-509B-104	Sequence 104, App
31	372.5	39.1	417	1	US-08-642-255-102	Sequence 102, App
32	372.5	39.1	417	1	US-08-707-237A-76	Sequence 76, App1
33	372.5	39.1	417	2	US-08-482-085B-104	Sequence 104, App
34	372.5	39.1	417	2	US-09-444-791A-104	Sequence 104, App
35	372.5	39.1	680	2	US-09-949-001-20	Sequence 5, App1
36	372.5	39.1	684	2	US-09-961-403-5	Sequence 4, App1
37	372.5	39.1	720	2	US-09-219-849-4	Sequence 4, App1
38	372.5	39.1	777	1	US-08-642-255-53	Sequence 53, App1
39	372.5	39.1	829	1	US-08-642-255-132	Sequence 132, App
40	372.5	39.1	829	1	US-08-397-633A-53	Sequence 68, App1
41	372.5	39.1	837	1	US-08-175-155-68	Sequence 103, App
42	372.5	39.1	837	1	US-08-477-509B-103	Sequence 101, App
43	372.5	39.1	837	1	US-08-642-255-101	Sequence 101, App
44	372.5	39.1	837	1	US-08-707-237A-75	Sequence 75, App1
45	372.5	39.1	837	2	US-08-482-085B-103	Sequence 103, App
46	372.5	39.1	837	2	US-09-444-791A-103	Sequence 50, App1
47	372.5	39.1	897	2	US-08-397-633A-50	Sequence 1, App1
48	372	39.1	1057	2	US-08-931-820-1	Sequence 16, App1
49	372	39.1	1057	2	US-10-153-469A-16	Sequence 20, App1
50	372	39.1	1057	2	US-10-153-469A-20	Sequence 16, App1
51	372	39.1	1057	2	US-10-104-889-16	Sequence 20, App1
52	372	39.1	1057	2	US-10-104-889-20	Sequence 11, App1
53	372	39.1	1107	2	US-10-153-469A-11	Sequence 8, App1
54	372	39.1	1107	2	US-10-104-889-11	Sequence 8, App1
55	372	39.1	1171	2	US-10-153-469A-8	Sequence 7065, Ap
56	372	39.1	1171	2	US-10-104-889-8	Sequence 10, App1
57	372	39.1	1218	2	US-09-949-016-7065	Sequence 10, App1
58	372	39.1	1388	2	US-10-153-469A-10	Sequence 15, App1
59	372	39.1	1388	2	US-10-104-889-10	Sequence 21, App1
60	372	39.1	1464	2	US-09-331-347-21	Sequence 15, App1
61	371.5	39.0	1680	2	US-09-949-001-15	Sequence 56, App1
62	371.5	39.0	1806	2	US-09-919-497-56	Sequence 10, App1
63	371	38.9	1017	2	US-08-468-996-10	Sequence 12, App1
64	370.5	38.9	1442	1	US-08-316-650-12	Sequence 6, App1
65	370.5	38.9	1442	4	PCT-US95-02251-12	Sequence 6, App1
66	370	38.9	1169	2	US-10-153-469A-6	Sequence 6, App1
67	370	38.9	1169	2	US-10-104-889-6	Sequence 6, App1
68	370	38.9	1341	2	US-08-363-825-18	Sequence 18, App1
69	370	38.9	1341	2	US-09-500-811-18	Sequence 18, App1
70	370	38.9	1341	2	US-09-570-573-18	Sequence 18, App1
71	370	38.9	1341	2	US-09-548-608-18	Sequence 18, App1
72	370	38.9	1461	2	US-09-585-887-9	Sequence 9, App1
73	370	38.9	1461	2	US-09-289-578-9	Sequence 9, App1
74	369.5	38.8	1060	2	US-08-531-820-3	Sequence 3, App1
75	369.5	38.8	1418	2	US-08-963-825-20	Sequence 20, App1
76	369.5	38.8	1418	2	US-09-010-999-1	Sequence 1, App1
77	369.5	38.8	1418	2	US-09-500-811-20	Sequence 20, App1
78	369.5	38.8	1418	2	US-09-570-573-20	Sequence 20, App1
79	369.5	38.8	1418	2	US-09-548-608-20	Sequence 20, App1
80	369	38.8	532	1	US-08-994-168-9	Sequence 9, App1
81	368	38.7	938	2	US-09-949-016-9992	Sequence 9992, Ap
82	366	38.4	1739	2	US-09-795-061-2	Sequence 2, App1
83	365.5	38.4	492	2	US-08-468-996-12	Sequence 12, App1
84	365.5	38.4	882	2	US-09-219-849-49	Sequence 49, App1
85	365	38.3	689	2	US-09-949-016-6137	Sequence 11276, A
86	362	38.0	595	2	US-09-819-849-48	Sequence 48, App1
87	362	38.0	595	2	US-09-819-849-50	Sequence 50, App1
88	361.5	38.0	1078	2	US-09-949-016-11185	Sequence 11185, A
89	361.5	38.0	1143	2	US-09-949-016-6137	Sequence 6137, Ap
90	360.5	37.9	347	2	US-09-823-497-1	Sequence 1, App1
91	360.5	37.9	623	2	US-09-829-348-3	Sequence 3, App1
92	360.5	37.9	626	2	US-09-829-348-12	Sequence 2, App1
93	360.5	37.9	684	1	US-08-555-669-12	Sequence 12, App1
94	360.5	37.9	684	2	US-09-073-663-1	Sequence 8209, Ap
95	360.5	37.9	1516	2	US-09-949-016-8209	Sequence 44, App1
96	359.5	37.8	219	2	US-10-153-469A-44	Sequence 46, App1
97	359.5	37.8	219	2	US-10-153-469A-46	Sequence 44, App1
98	359.5	37.8	219	2	US-10-104-889-44	Sequence 46, App1
99	359.5	37.8	219	2	US-10-104-889-46	Sequence 46, App1
100	359.5	37.8	234	2	US-09-895-674A-1	Sequence 1, App1

## ALIGNMENTS

```
RESULT 1
US-10-000-489-58
; Sequence 58, Application US/10000489
; Patent No. 6794363
; GENERAL INFORMATION:
; APPLICANT: Benjann, Stephanie
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US6.DIV
; CURRENT APPLICATION NUMBER: US/10/000,489
; CURRENT FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: JPatent
; SEQ ID NO 58
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-000-489-58

Query Match      100.0%; Score 952; DB 2; Length 163;
Best Local Similarity 100.0%; Pred. No. 1,1e-65;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPPGKGTGHPGLPGPKDCKGPPPGSTGRPGAEGEPGAMGPGRPGPPGHVGP 60
DB 1 MGPPGKGTGHPGLPGPKDCKGPPPGSTGRPGAEGEPGAMGPGRPGPPGHVGP 60
QY 61 PGPAGISAVGLKGDRTGATGERGLAGLPGQGPDPGPGPGPGYGMGATGPMGQGGIFG 120
DB 61 PGPAGISAVGLKGDRTGATGERGLAGLPGQGPDPGPGPGPGYGMGATGPMGQGGIFG 120
QY 121 IQPPPGMGQPGKAGHCNPSDCGAMPMEQYPPMKTMGPPFG 163
DB 121 IQPPPGMGQPGKAGHCNPSDCGAMPMEQYPPMKTMGPPFG 163

RESULT 2
US-09-949-016-6136
; Sequence 6136, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6136
```

```
LENGTH: 1603
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6136

Query Match      88.3%; Score 840.5; DB 2; Length 1603;
Best Local Similarity 44.5%; Pred. No. 2.9e-56;
Matches 163; Conservative 0; Mismatches 0; Indels 203; Gaps 1;

QY 1 MGPPGKGTGHPGLPGPKDCKGPPPGSTGRPGAEGEPGAMGPGRPGPPGHVGP 60
DB 1238 MGPPGKGTGHPGLPGPKDCKGPPPGSTGRPGAEGEPGAMGPGRPGPPGHVGP 1297
QY 61 PGPAGISAVGLKGDRTGATGERGLAGLPGQ----- 92
DB 1298 PGPAGISAVGLKGDRTGATGERGLAGLPGQPPGHPGPPGPGRTDGAAGKGGPPGKQ 1357
QY 93 ----- 92
DB 1358 GPPGPPKDPGAAQKQAGEKRGAMGPGPGKSGSMGVGPPGPPAGERGHPGAPPS 1417
QY 93 ----- 92
DB 1418 GSPGLPGVPSGMDWNTYDEIRFLRQELIKMFDERMAVYTSRMQPFMEMAALPGRPP 1477
QY 93 -----PGPPG 97
DB 1478 GKDGAAPRGAPSPGLPGQIGREGROGLPGVRLPGTKGEXKDIGIGIAGENGLPGPPG 1537
QY 98 PGPAGYGMGATGPMGQGGIFGIPGPPGPMGPGKAGHCNPSDCGAMPMEQYPPMKT 157
DB 1538 PGPAGYGMGATGPMGQGGIFGIPGPPGPMGPGKAGHCNPSDCGAMPMEQYPPMKT 1597
QY 158 MKGPPG 163
DB 1598 MKGPPG 1603

RESULT 3
US-09-949-016-10910
; Sequence 10910, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10910
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10910

Query Match      88.3%; Score 840.5; DB 2; Length 1603;
Best Local Similarity 44.5%; Pred. No. 2.9e-56;
Matches 163; Conservative 0; Mismatches 0; Indels 203; Gaps 1;

QY 1 MGPPGKGTGHPGLPGPKDCKGPPPGSTGRPGAEGEPGAMGPGRPGPPGHVGP 60
DB 1244 MGPPGKGTGHPGLPGPKDCKGPPPGSTGRPGAEGEPGAMGPGRPGPPGHVGP 1303
QY 61 PGPAGISAVGLKGDRTGATGERGLAGLPGQ----- 92
```

Db 1304 PGQPGAGISAVGLKGDRTGATGERGLAGLPQPGPPGHPPGPBPGPTDGAAGKEGPPKQ 1363  
QY 93 ----- 92  
Db 1364 GFYGPQPKDPAAGQKGAAGEKGRAGMPGPGKSGSMKPVGPBPAGRGHFGARGPS 1423  
QY 93 ----- 92  
Db 1424 GSPFLPGVPSMGDMVYDEIKRFIRQELIKMFDERMAVYYSNMQFPMEMAAAPGRPP 1483  
QY 93 -----RPPG 97  
Db 1484 GKDAAPRPGAPSSPGLPGQIGRBGRQLPGVHGLPCTKGEKDDIGIAGENGGLPQPPG 1543  
QY 98 PGPPGYGKMGATGPMQOGIPGIPPPPMQPGKAGHCNPSDCFGAMPEDQYPPMK 157  
Db 1544 PGPPGYGKMGATGPMQOGIPGIPPPPMQPGKAGHCNPSDCFGAMPEDQYPPMK 1603  
QY 158 MKGPPG 163  
Db 1604 MKGPPG 1609

RESULT 4  
US-08-642-255-32  
; Sequence 32, Application US/08642255  
; Patent No. 5773249  
; GENERAL INFORMATION:  
; APPLICANT: CARPELLO, Joseph  
; APPLICANT: FERRARI, Franco A.  
; TITLE OF INVENTION: High Molecular Weight Collagen-Like  
; TITLE OF INVENTION: Protein Polymers  
; NUMBER OF SEQUENCES: 135  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/642,255  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: ROWLAND, Berttram I.  
; REGISTRATION NUMBER: 20,015  
; REFERENCE/DOCKET NUMBER: A5556-3/BIR  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 494-8700  
; TELEFAX: (415) 494-8771  
; TELEX: 910 277299 FHT UR  
; INFORMATION FOR SEQ ID NO: 32:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 330 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-642-255-32

Query Match 44.0%; Score 419; DB 1; Length 330;  
Best Local Similarity 54.2%; Pred. No. 8,4e-25;  
Matches 83; Conservative 2; Mismatches 60; Indels 8; Gaps 2;  
QY 2 GPPGFGKGTGHPGLPGPKDDCGKPPPGSTGRPGABEGPGAMPQGRPPGPHVGPGRPP 61  
Db 157 GPPGPPGPPGAPGPPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 216

QY 62 GQPGPAG-ISAVGLKGDRTGATGERGLAGLPQPGPPPGQPPG-----YGMGATGPM 113  
Db 217 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 276  
QY 114 GQGGIPGIPGPPPMQPGKAGHCNPSDCFGAM 146  
Db 277 GPPGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 309

RESULT 5  
US-07-609-716-65  
; Sequence 65, Application US/07609716  
; Patent No. 5514581  
; GENERAL INFORMATION:  
; APPLICANT: Ferrari, Franco A.  
; APPLICANT: Carpello, Joseph  
; TITLE OF INVENTION: Functional Recombinantly Prepared  
; TITLE OF INVENTION: Synthetic Protein Polymer  
; NUMBER OF SEQUENCES: 118  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: US  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/609,716  
; FILING DATE: 06-NOV-1990  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rowland, Berttram I.  
; REGISTRATION NUMBER: 20015  
; REFERENCE/DOCKET NUMBER: A-55186-3/BIR  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-781-1989  
; TELEFAX: 415-398-3249  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 408 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-609-716-65

Query Match 44.0%; Score 419; DB 1; Length 408;  
Best Local Similarity 54.2%; Pred. No. 1e-24;  
Matches 83; Conservative 2; Mismatches 60; Indels 8; Gaps 2;  
QY 2 GPPGFGKGTGHPGLPGPKDDCGKPPPGSTGRPGABEGPGAMPQGRPPGPHVGPGRPP 61  
Db 235 GPPGPPGPPGAPGPPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 294  
QY 62 GQPGPAG-ISAVGLKGDRTGATGERGLAGLPQPGPPPGQPPG-----YGMGATGPM 113  
Db 295 GPPGPPGPPGAPGPPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 354  
QY 114 GQGGIPGIPGPPPMQPGKAGHCNPSDCFGAM 146  
Db 355 GPPGPPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 387  
RESULT 6  
US-08-475-411A-65  
; Sequence 65, Application US/08475411A  
; Patent No. 6140072

```

; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,411A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccatin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-9/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-475-411A-65

Query Match      44.0%; Score 419; DB 2; Length 408;
Best Local Similarity 54.2%; Pred. No. 1e-24;
Matches 83; Conservative 2; Mismatches 60; Indels 8; Gaps 2;

QY 2 GPPGFGKGTGHPGLPKPKDCKPSPGSTGRPGAEGSGAMPGQRPDPGHVGPDP 61
   |||||
DB 235 GPPGPPGPPGAGPPPPPPPPGPPGAGVSGAPGPPGPPPPGPPGAPGPPGPP 294
   |||||
QY 62 GPPGPPG-ISAVALKGDRCATGERGLAGLPGQPPGPPGPPG-----YKMGATGPM 113
   |||||
DB 295 GPPGPPGPPGPPGPPPPPPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 354
   |||||
QY 114 GQGGIPGIPGPPGPPMGQPKAGHCNPSDFGAM 146
   |||||
DB 355 GPPGPPGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 387
   |||||

RESULT 7
US-08-478-029A-65
; Sequence 65, Application US/08478029A
; Patent No. 6184348
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
```

```

; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,029A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccatin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-478-029A-65

Query Match      44.0%; Score 419; DB 2; Length 408;
Best Local Similarity 54.2%; Pred. No. 1e-24;
Matches 83; Conservative 2; Mismatches 60; Indels 8; Gaps 2;

QY 2 GPPGFGKGTGHPGLPKPKDCKPSPGSTGRPGAEGSGAMPGQRPDPGHVGPDP 61
   |||||
DB 235 GPPGPPGPPGAGPPPPPPPPGPPGAGVSGAPGPPGPPPPGPPGAPGPPGPP 294
   |||||
QY 62 GPPGPPG-ISAVALKGDRCATGERGLAGLPGQPPGPPGPPG-----YKMGATGPM 113
   |||||
DB 295 GPPGPPGPPGPPGPPPPPPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 354
   |||||
QY 114 GQGGIPGIPGPPGPPMGQPKAGHCNPSDFGAM 146
   |||||
DB 355 GPPGPPGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 387
   |||||

RESULT 8
US-07-609-716-66
; Sequence 66, Application US/07609716
; Patent No. 5514581
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
```

```

? TITLE OF INVENTION: Synthetic Protein Polymer
? NUMBER OF SEQUENCES: 118
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Flehr, Honbach, Test, Albritten & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/609, 716
FILING DATE: 06-NOV-1990
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: A-55186-3/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEX: 415-398-3249
INFORMATION FOR SEO ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 357 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-609-716-66

Query Match      43.3% Score 412.5 DB 1 Length 357;
Best Local Similarity 42.7%; Pred No. 2,8e-24;
Matches 90; Conservative 6; Mismatches 54; Indels 61; Gaps 6

QY    2 GPFGFKGTGHPGLPGPKDGGKPPGPSTGNP-----GAEGEPGANGPQGRRPGPHGV 55
Db     ||||| :||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY    52 GPPRPPGPGRPLGPRKGDRGDAGPKADSPGAPGVSGAAGPPGPPEPPPCAPAR 111
Db     ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY    56 GPPRPQPQPGAGISAVGLKDGRGATGERGL-----AG 88
Db     ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY   112 GPPRPPGPGRGPLP--GPKGDRGDAGPKADSGAPAGPVGS PGAPGPBPBPPBPPBG 169
Db     ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY    89 LPQGPGPPRGPPGPCGY-----GKKGAGCPMGQQGILPIPRPGPMGQPG 132
Db     ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY   170 APGPGRPGRPPGPRLPGPKCDRDGDAGPKADSGAPGVPVSGAPGPBPBPPBPPBG 229
Db     ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY   133 KAGHCNPSDCGAMPMEQQYPMTKMKGPF 163
Db     |       :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY   230 APGPGRPPGPBG-----PP--GIKPKG 250
Db     |       :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

RESULT 9
US-08-642-255-33
Sequence 33, Application US/08642255
Patent No. 5773249
GENERAL INFORMATION:
APPLICANT: CAPPELLO, Joseph
APPLICATOR: FERRARI, Franco A.
TITLE OF INVENTION: High Molecular Weight Collagen-Like
TITLE OF INVENTION: Protein Polymers
NUMBER OF SEQUENCES: 135
CORRESPONDENCE ADDRESS:
ADDRESSEES: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
```

```

: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/642,255
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: ROWLAND, Bertram I.
: REGISTRATION NUMBER: 20,015
: REFERENCE/DOCKET NUMBER: A5556-3/BIR
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 494-8700
: TELEFAX: (415) 494-8771
: TELEX: 910 277299 FHT UR
: INFORMATION FOR SEQ ID NO: 33:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 357 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-642-255-33

Query Match 43.3%; Score 412.5; DB 1; Length 357;
Best Local Similarity 42.7%; Pred. No. 2.8e-24;
Matches 90; Conservative 6; Mismatches 54; Indels 61; Gaps 6;

Qy 2 GPPGFKKGTGHPGLGPKKGDGCKRGCPGSGTGRP-----GAEGRGAMGPGGRPGPGPHV 55
Db 52 GPPPPGPPGPPGGLPPPGKGDGADPKGADGSPGPPAGPVGSPGAPGPPGPPGPPGAP 111
Qy 56 GPPPPGQOPGAGISAVGLKGDRCATGTERGL-----AG 88
Db 112 GPPPPGPPGPPGPPGLP--GPKDGRDADGPKGADGSPGAPGVGSPGAPGPPPPGPPGPPG 169
Qy 89 LPQGPPGPPGPPGPPGY-----GKMGATGPMGQGIPIGPPGPMGQPG 132
Db 170 APGPPGPPGPPGPPGPPGLPKPKKDRDADGPKGADGSPGAPGVGSPGAPGPPGPPGPPG 229
Qy 133 KAGHCNPSDCRGAMPMEQYIPPMKTMKPGPG 163
Db 230 APGPPGPPGPPG-----PP--GLDGPKG 250

RESULT 10
US-08-475-411A-66
: Sequence 66, Application US/08475411A
: Patent No. 6140072
: GENERAL INFORMATION:
: APPLICANT: Ferrari, Franco A.
: TITLE OF INVENTION: Functional Recombinantly Prepared
: TITLE OF INVENTION: Synthetic Protein Polymer
: NUMBER OF SEQUENCES: 119
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Flehr, Hobbach, Teet, Albritton & Herbert
: STREET: Four Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: CA
: COUNTRY: US
: ZIP: 94111
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/475,411A
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:

```

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/609,716  
FILING DATE: 06-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/269,429  
FILING DATE: 09-NOV-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/114,618  
FILING DATE: 29-OCT-1987  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 06/927,258  
FILING DATE: 04-NOV-1986  
ATTORNEY/AGENT INFORMATION:  
NAME: Treccartin, Richard F.  
REGISTRATION NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-781-1989  
TELEFAX: 415-398-3249  
INFORMATION FOR SEQ ID NO: 66:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 357 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein

Query Match	Similarity	43.3%	Score 412.5	DB 2	Length 357
Best Local	42.7%	Pred. No.28-24			
Matches	90	Conservative 6	Matches 54	Indels 61	Gaps 6
Qy	2	GPFGKGTGTHPLPPGKGDGCGKPGPGSTGRP-----GAEGEPGAMGQGRPPGPHV	55		
Db	52	GPFPFPFPFPFPGLPGRFKDGRGDAGPKGADGSPGDPAGVSGPAGPFPFPFPFPFP	111		
Qy	56	GPFPFPFPFPAGTISAVGLKGRATGATGRGL-----AG	88		
Db	112	GPFPFPFPFPFPGLP- -GPKGDRGDAGPKGADGSPGAPGVGSPGAPFPFPFPFPFP	169		
Qy	89	LPQGPFPFPFPQPPGPGY-----GKMGATGPMGQOQIPGIPGPPGPMGQPG	132		
Db	170	APGPFPFPFPFPFPGLPFGKDRDRDAGKGAAGSPGAPGVGSPGAPFPFPFPFPFP	229		
Qy	133	KAGHCNPSDCGAMPMEQVPEPMKTMGKPGF	163		
Db	230	APGPFPFPFPFP-----PP- -GLPFGK	250		

RESULT 12  
 US-08-494-168-10  
 : Sequence 10, Application US/08494168  
 : Patent No. 5731192  
 : GENERAL INFORMATION:  
 : APPLICANT: Reeders, Stephen T.  
 : APPLICANT: Zhou, Jing  
 : TITLE OF INVENTION: Collagen COL4A6: Gene, Protein and Method  
 : TITLE OF INVENTION: of Detecting Collagen Deficiency  
 : NUMBER OF SEQUENCES: 10  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Foley & Lardner  
 : STREET: 3000 K Street, N.W., Suite 500  
 : CITY: Washington, D.C.  
 : COUNTRY: USA  
 : ZIP: 20007-5109  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: PatentIn Release #1.0, Version #1.25  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/494.168  
 :





QY	161	PF	163
Db	290	PAG	292

RESULT 15  
US-09-219

US-09-219-849-3

; Sequence 3, Application US/09219849  
; patent No. 6150081

GENERAL INFORMATION:

APPLICANT: VAN HEERDE, GEORGE V.

APPLICANT: VAN RIJN, ALEXIS C.  
BOUWSTRA, JAN B.

APPLICANT: BOUMSIKA, JAN B.  
DE WOLF, FREDERBT

APPLICANT: MOOBROEK, ANDREAS

APPLICANT: WERTEN, MARC W.T.

APPLICANT: WIND, RICHELE D.

APPLICANT: VAN DEN BOSCH, TANJA J.

TITLE OF INVENTION: SILVER HALIDE

TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE METHOD OF TREATMENT: PREPARATION THEREOF

TITLE OF INVENTION: PREPARATION THEREOF  
; TITLE REFERENCE: 3728-2

FILE REFERENCE: 2148-2

CURRENT FILING DATE: 1998-12-23

NUMBER OF SEQ ID NOS: 50

; SOFTWARE: PatentIn Ver.

; SEQ ID NO 3

LENGTH: 504

TYPE: PRI  
ORGANISM: A

FEATURE: ALLICIN sequence

OTHER INFORMATION: Description

OTHER INFORMATION: amino acid

US-09-219-849-3

1

Query Match	40.8%;	Score 388.5;	DB 2;	Length 504;
Best 1000 Similar	47.5%;	Prod NO 2 50-33;		

Matches 87: Conservative 2: Mismatches 71  
Best Local Similarity 47.36; Pred. NO. 2.0e-22;

MACCIB	01	COMBETVCLVC	12	INDMNCUCC	13	201	202

[illegible]

Search completed: December 1, 2005, 08:28:05  
Job time : 53 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: December 1, 2005, 08:27:13 ; Search time 162 Seconds  
(without alignments)  
420.409 Million cell updates/sec

Title: US-09-924-340-58

Perfect score: 952

Sequence: 1 MGPMPGKGTGHTGHPGPKG.....GAMPMEQYPMKTMKPGPG 163

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

Published Applications\_AA\_Main:\*  
1: /cgn2\_6/prodata/1/pubppa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/prodata/1/pubppa/US08\_PUBCOMB.pep:\*  
3: /cgn2\_6/prodata/1/pubppa/US09\_PUBCOMB.pep:\*  
4: /cgn2\_6/prodata/1/pubppa/US10\_PUBCOMB.pep:\*  
5: /cgn2\_6/prodata/1/pubppa/US10B\_PUBCOMB.pep:\*  
6: /cgn2\_6/prodata/1/pubppa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	952	100.0	163	3	US-09-992-600A-58	Sequence 58, App1
2	952	100.0	163	3	US-09-924-340-58	Sequence 58, App1
3	952	100.0	163	3	US-09-992-095B-58	Sequence 58, App1
4	952	100.0	163	3	US-09-999-570-58	Sequence 58, App1
5	952	100.0	163	4	US-10-000-489-58	Sequence 58, App1
6	952	100.0	163	4	US-10-000-986-58	Sequence 58, App1
7	952	100.0	163	4	US-10-154-678-58	Sequence 58, App1
8	952	100.0	163	5	US-10-838-854-58	Sequence 58, App1
9	840.5	88.3	1603	5	US-10-723-860-4225	Sequence 39, App1
10	405	42.5	180	5	US-10-797-606-39	Sequence 218, App
11	395	41.5	67	4	US-10-001-843-218	Sequence 218, App
12	395	41.5	67	6	US-11-005-609-218	Sequence 2221, App
13	390	41.0	910	5	US-10-723-860-2221	Sequence 2221, App
14	390	41.0	1685	5	US-10-450-763-35198	Sequence 45978, A
15	390	41.0	1685	5	US-10-450-763-45978	Sequence 45978, A
16	388.5	40.8	504	4	US-10-342-331-3	Sequence 3, App1
17	385	40.4	1487	4	US-10-157-031-128	Sequence 128, App
18	384	40.3	1466	4	US-10-402-089-4	Sequence 4, App1
19	384	40.3	1466	4	US-10-402-089-6	Sequence 4, App1
20	384	40.3	1466	4	US-10-402-072A-4	Sequence 6, App1
21	384	40.3	1466	4	US-10-402-072A-6	Sequence 6, App1
22	383.5	40.3	1838	5	US-10-852-335A-162	Sequence 162, App
23	383.5	40.3	1838	5	US-10-287-436A-464	Sequence 162, App
24	383.5	40.3	1838	5	US-10-287-436A-1161	Sequence 1161, App
25	382	40.1	1266	4	US-10-367-094-85	Sequence 85, App1
26	382	40.1	1266	4	US-10-367-094-87	Sequence 87, App1
27	382	40.1	1679	4	US-10-367-094-89	Sequence 89, App1

28	381.5	40.1	703	4	US-10-219-449-4	Sequence 4, App1
29	381.5	40.1	703	5	US-10-958-858-4	Sequence 4, App1
30	381.5	40.1	717	4	US-10-219-449-2	Sequence 2, App1
31	381.5	40.1	717	5	US-10-958-858-2	Sequence 2, App1
32	381.5	40.1	733	4	US-10-138-588-40	Sequence 40, App1
33	381.5	40.1	840	4	US-10-567-094-82	Sequence 82, App1
34	380.5	40.0	638	4	US-10-138-588-42	Sequence 42, App1
35	380	39.9	1767	5	US-10-741-600-1476	Sequence 1476, App
36	380	39.9	1767	5	US-10-741-600-1478	Sequence 1478, App
37	380	39.9	1806	5	US-10-741-600-1479	Sequence 1479, App
38	380	39.9	1818	5	US-10-741-600-1475	Sequence 1475, App
39	380	39.9	1818	5	US-10-741-600-1477	Sequence 1477, App
40	378.5	39.8	638	4	US-10-001-887-108	Sequence 108, App
41	378.5	39.8	638	5	US-10-958-863-108	Sequence 108, App
42	378.5	39.8	1745	3	US-09-795-061-4	Sequence 4, App1
43	378.5	39.8	1745	5	US-10-723-860-2660	Sequence 2660, App
44	378.5	39.7	1466	4	US-10-402-089-12	Sequence 12, App1
45	377.5	39.7	1466	4	US-10-402-072A-12	Sequence 12, App1
46	377.5	39.7	1466	4	US-10-058-124-21	Sequence 21, App1
47	377	39.6	1449	4	US-10-402-089-8	Sequence 8, App1
48	377	39.6	1449	4	US-10-402-072A-8	Sequence 8, App1
49	377	39.6	1466	3	US-09-918-715-226	Sequence 226, App
50	377	39.6	1466	3	US-10-177-293-68	Sequence 68, App1
51	377	39.6	1466	4	US-10-301-822-33	Sequence 33, App1
52	377	39.6	1466	4	US-10-257-021-72	Sequence 72, App1
53	377	39.6	1466	4	US-10-357-851-3	Sequence 3, App1
54	377	39.6	1466	4	US-10-358-024-3	Sequence 3, App1
55	377	39.6	1466	4	US-10-734-564-103	Sequence 103, App
56	377	39.6	1466	4	US-10-474-794-226	Sequence 226, App
57	377	39.6	1466	5	US-10-852-335A-159	Sequence 159, App
58	377	39.6	1466	5	US-10-979-159-226	Sequence 226, App
59	377	39.6	1466	5	US-10-287-436A-494	Sequence 494, App
60	377	39.6	1466	5	US-10-287-436A-1151	Sequence 1151, App
61	377	39.6	1466	5	US-10-287-436A-1189	Sequence 1189, App
62	377	39.6	1469	5	US-10-450-763-45550	Sequence 45550, A
63	377	39.3	1518	5	US-10-450-763-53038	Sequence 53038, A
64	374	39.3	544	5	US-10-658-989A-4	Sequence 4, App1
65	373	39.2	1463	4	US-10-402-089-2	Sequence 2, App1
66	373	39.2	1463	4	US-10-402-072A-2	Sequence 2, App1
67	373	39.1	202	4	US-10-029-386-32935	Sequence 32935, A
68	372.5	39.1	417	5	US-10-096-986-104	Sequence 104, App
69	372.5	39.1	528	5	US-10-719-993-512	Sequence 512, App
70	372.5	39.1	550	5	US-10-719-993-498	Sequence 498, App
71	372.5	39.1	680	4	US-10-295-027-110	Sequence 110, App
72	372.5	39.1	680	5	US-10-643-795A-118	Sequence 118, App
73	372.5	39.1	680	5	US-10-948-518-118	Sequence 118, App
74	372.5	39.1	684	3	US-09-961-403-5	Sequence 5, App1
75	372.5	39.1	684	5	US-10-719-993-502	Sequence 502, App
76	372.5	39.1	720	4	US-10-342-331-4	Sequence 4, App1
77	372.5	39.1	837	4	US-10-096-986-103	Sequence 103, App
78	372.5	39.1	837	4	US-10-332-175-27	Sequence 27, App1
79	372.5	39.1	333	6	US-11-139-377-27	Sequence 27, App1
80	372	39.1	333	6	US-10-332-175-26	Sequence 26, App1
81	372	39.1	510	6	US-10-332-175-26	Sequence 26, App1
82	372	39.1	510	6	US-10-332-175-26	Sequence 26, App1
83	372	39.1	660	5	US-10-901-816A-3	Sequence 3, App1
84	372	39.1	660	5	US-10-901-816A-4	Sequence 4, App1
85	372	39.1	662	5	US-10-301-816A-9	Sequence 9, App1
86	372	39.1	662	5	US-10-301-816A-10	Sequence 10, App1
87	372	39.1	1014	5	US-10-901-816A-11	Sequence 11, App1
88	372	39.1	1014	5	US-10-901-816A-12	Sequence 12, App1
89	372	39.1	1014	5	US-10-901-816A-13	Sequence 13, App1
90	372	39.1	1014	5	US-10-901-816A-7	Sequence 7, App1
91	372	39.1	1014	5	US-10-901-816A-8	Sequence 8, App1
92	372	39.1	1014	5	US-10-901-816A-9	Sequence 9, App1
93	372	39.1	1014	5	US-10-901-816A-10	Sequence 10, App1
94	372	39.1	1014	5	US-10-901-816A-11	Sequence 11, App1
95	372	39.1	1014	5	US-10-901-816A-12	Sequence 12, App1
96	372	39.1	1014	5	US-10-901-816A-13	Sequence 13, App1
97	372	39.1	1057	4	US-10-104-889-16	Sequence 16, App1
98	372	39.1	1057	4	US-10-104-889-20	Sequence 20, App1
99	372	39.1	1057	5	US-10-104-793-16	Sequence 16, App1
100	372	39.1	1057	5	US-10-104-793-20	Sequence 20, App1

## ALIGNMENTS

## RESULT 1

US-09-992-600A-58

; Sequence 58, Application US/09992600A

; Publication No. US20030027161A1

; GENERAL INFORMATION:

; APPLICANT: Benjamin, Stephanie

; APPLICANT: Tanaka, Hiroaki

; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF

; FILE REFERENCE: 91.US4.DIV

; CURRENT APPLICATION NUMBER: US/09/992,600A

; PRIOR FILING DATE: 2001-11-13

; PRIOR APPLICATION NUMBER: US 09/924,340

; PRIOR FILING DATE: 2001-08-06

; PRIOR APPLICATION NUMBER: PCT/IB01/01715

; PRIOR FILING DATE: 2001-08-06

; PRIOR APPLICATION NUMBER: US 60/305,456

; PRIOR FILING DATE: 2001-07-13

; PRIOR APPLICATION NUMBER: US 60/302,277

; PRIOR FILING DATE: 2001-06-29

; PRIOR APPLICATION NUMBER: US 60/298,698

; PRIOR FILING DATE: 2001-06-15

; PRIOR APPLICATION NUMBER: US 60/293,574

; PRIOR FILING DATE: 2001-05-25

; NUMBER OF SEQ ID NOS: 114

; SOFTWARE: JPatent

; SEQ ID NO 58

; LENGTH: 163

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-992-600A-58

Query Match 100.0%; Score 952; DB 3; Length 163;

Best Local Similarity 100.0%; Pred. No. 7.4e-54;

Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MGPPGKGTGTHGGLPGPKDCKPGPPGSTRPGABEGPGAMGPGRPPGHVPPPP 60
DB 1 MGPPGKGTGTHGGLPGPKDCKPGPPGSTRPGABEGPGAMGPGRPPGHVPPPP 60
QY 61 PGQPGAGISAVGLKGDRTATGERGLAGLPGQPGPPGPPGPGYGMGATGPMGQGGIRG 120
DB 61 PGQPGAGISAVGLKGDRTATGERGLAGLPGQPGPPGPPGPGYGMGATGPMGQGGIRG 120
QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMMGPFG 163
DB 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMMGPFG 163
```

## RESULT 2

US-09-924-340-58

; Sequence 58, Application US/09924340

; Publication No. US20030027248A1

; GENERAL INFORMATION:

; APPLICANT: Benjamin, Stephanie

; APPLICANT: Tanaka, Hiroaki

; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF

; FILE REFERENCE: 91.US2.REG

; CURRENT APPLICATION NUMBER: US/09/924,340

; PRIOR FILING DATE: 2001-08-06

; PRIOR APPLICATION NUMBER: US 60/305,456

; PRIOR FILING DATE: 2001-07-13

; PRIOR APPLICATION NUMBER: US 60/302,277

; PRIOR FILING DATE: 2001-06-29

; PRIOR APPLICATION NUMBER: US 60/298,698

; PRIOR FILING DATE: 2001-06-15

; PRIOR APPLICATION NUMBER: US 60/293,574

; PRIOR FILING DATE: 2001-05-25

; NUMBER OF SEQ ID NOS: 112

```
; SOFTWARE: JPatent
; SEQ ID NO 58
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-924-340-58
```

```
Query Match 100.0%; Score 952; DB 3; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.4e-54;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MGPPGKGTGTHGGLPGPKDCKPGPPGSTRPGABEGPGAMGPGRPPGHVPPPP 60
DB 1 MGPPGKGTGTHGGLPGPKDCKPGPPGSTRPGABEGPGAMGPGRPPGHVPPPP 60
QY 61 PGQPGAGISAVGLKGDRTATGERGLAGLPGQPGPPGPPGPGYGMGATGPMGQGGIRG 120
DB 61 PGQPGAGISAVGLKGDRTATGERGLAGLPGQPGPPGPPGPGYGMGATGPMGQGGIRG 120
QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMMGPFG 163
DB 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMMGPFG 163
```

## RESULT 3

US-09-992-095B-58

; Sequence 58, Application US/09992095B

; Publication No. US20030157485A1

; GENERAL INFORMATION:

; APPLICANT: Benjamin, Stephanie

; APPLICANT: Tanaka, Hiroaki

; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF

; FILE REFERENCE: 91.US5.DIV

; CURRENT APPLICATION NUMBER: US/09/992,095B

; PRIOR FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: US 09/924,340

; PRIOR FILING DATE: 2001-08-06

; PRIOR APPLICATION NUMBER: PCT/IB01/01715

; PRIOR FILING DATE: 2001-08-06

; PRIOR APPLICATION NUMBER: US 60/305,456

; PRIOR FILING DATE: 2001-07-13

; PRIOR APPLICATION NUMBER: US 60/302,277

; PRIOR FILING DATE: 2001-06-29

; PRIOR APPLICATION NUMBER: US 60/298,698

; PRIOR FILING DATE: 2001-06-15

; PRIOR APPLICATION NUMBER: US 60/293,574

; PRIOR FILING DATE: 2001-05-25

; NUMBER OF SEQ ID NOS: 112

; SOFTWARE: JPatent

; SEQ ID NO 58

; LENGTH: 163

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-992-095B-58

```
Query Match 100.0%; Score 952; DB 3; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.4e-54;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MGPPGKGTGTHGGLPGPKDCKPGPPGSTRPGABEGPGAMGPGRPPGHVPPPP 60
DB 1 MGPPGKGTGTHGGLPGPKDCKPGPPGSTRPGABEGPGAMGPGRPPGHVPPPP 60
QY 61 PGQPGAGISAVGLKGDRTATGERGLAGLPGQPGPPGPPGPGYGMGATGPMGQGGIRG 120
DB 61 PGQPGAGISAVGLKGDRTATGERGLAGLPGQPGPPGPPGPGYGMGATGPMGQGGIRG 120
QY 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMMGPFG 163
DB 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMMGPFG 163
```

## RESULT 4

US-09-999-570-58  
; Sequence 58, Application US/09999570  
; Publication No. US20030170628A1  
; GENERAL INFORMATION:  
; APPLICANT: Benjamin, Stephane  
; APPLICANT: Tanaka, Hiroaki  
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF  
; FILE REFERENCE: G-091US08DIY  
; CURRENT APPLICATION NUMBER: US/09/999,570  
; PRIOR FILING DATE: 2001-06-14,340  
; PRIOR APPLICATION NUMBER: US 09/924,340  
; PRIOR FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: PCT/IB01/01715  
; PRIOR FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: US 60/305,456  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/302,277  
; PRIOR FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/298,698  
; PRIOR FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: US 60/293,574  
; PRIOR FILING DATE: 2001-05-25  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: JPatent  
; SEQ ID NO 58  
; LENGTH: 163  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-999-570-58

Query Match 100.0%; Score 952; DB 3; Length 163;  
Best Local Similarity 100.0%; Pred. No. 7,4e-54;  
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPPGKGTGTHGLGPKPKDCCGPPPGSTGTPGAGGEGGANGPQGRPPGHVGP 60  
DB 1 MGPPGKGTGTHGLGPKPKDCCGPPPGSTGTPGAGGEGGANGPQGRPPGHVGP 60  
QY 61 PGQGPAGISAVGLKDDRGATGGERGLAGLPQGPQPPGPGPGYGMKATGPMGQOQIPG 120  
DB 61 PGQGPAGISAVGLKDDRGATGGERGLAGLPQGPQPPGPGPGYGMKATGPMGQOQIPG 120  
QY 121 IPGPPGPMGQPGKAGHCNPSCFCGAMPMEQOYPPMTKMGPF 163  
DB 121 IPGPPGPMGQPGKAGHCNPSCFCGAMPMEQOYPPMTKMGPF 163

RESULT 5  
US-10-000-489-58  
; Sequence 58, Application US/10000489  
; Publication No. US2003092011A1  
; GENERAL INFORMATION:  
; APPLICANT: Benjamin, Stephane  
; APPLICANT: Tanaka, Hiroaki  
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF  
; FILE REFERENCE: 91.US6.DIV  
; CURRENT APPLICATION NUMBER: US/10/000,489  
; PRIOR FILING DATE: 2001-11-14,340  
; PRIOR APPLICATION NUMBER: US 09/924,340  
; PRIOR FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: PCT/IB01/01715  
; PRIOR FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: US 60/305,456  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/302,277  
; PRIOR FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/298,698  
; PRIOR FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: US 60/293,574  
; PRIOR FILING DATE: 2001-05-25  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: JPatent  
; SEQ ID NO 58

; LENGTH: 163  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-000-489-58

Query Match 100.0%; Score 952; DB 4; Length 163;  
Best Local Similarity 100.0%; Pred. No. 7,4e-54;  
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPPGKGTGTHGLGPKPKDCCGPPPGSTGTPGAGGEGGANGPQGRPPGHVGP 60  
DB 1 MGPPGKGTGTHGLGPKPKDCCGPPPGSTGTPGAGGEGGANGPQGRPPGHVGP 60  
QY 61 PGQGPAGISAVGLKDDRGATGGERGLAGLPQGPQPPGPGPGYGMKATGPMGQOQIPG 120  
DB 61 PGQGPAGISAVGLKDDRGATGGERGLAGLPQGPQPPGPGPGYGMKATGPMGQOQIPG 120  
QY 121 IPGPPGPMGQPGKAGHCNPSCFCGAMPMEQOYPPMTKMGPF 163  
DB 121 IPGPPGPMGQPGKAGHCNPSCFCGAMPMEQOYPPMTKMGPF 163

RESULT 6  
US-10-000-986-58  
; Sequence 58, Application US/10000986  
; Publication No. US2003096247A1  
; GENERAL INFORMATION:  
; APPLICANT: Benjamin, Stephane  
; APPLICANT: Tanaka, Hiroaki  
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF  
; FILE REFERENCE: 91.US9.DIV  
; CURRENT APPLICATION NUMBER: US/10/000,986  
; PRIOR FILING DATE: 2001-11-14,340  
; PRIOR APPLICATION NUMBER: US 09/924,340  
; PRIOR FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: PCT/IB01/01715  
; PRIOR FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: US 60/305,456  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/302,277  
; PRIOR FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/298,698  
; PRIOR FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: US 60/293,574  
; PRIOR FILING DATE: 2001-05-25  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: JPatent  
; SEQ ID NO 58  
; LENGTH: 163  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-000-986-58

Query Match 100.0%; Score 952; DB 4; Length 163;  
Best Local Similarity 100.0%; Pred. No. 7,4e-54;  
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPPGKGTGTHGLGPKPKDCCGPPPGSTGTPGAGGEGGANGPQGRPPGHVGP 60  
DB 1 MGPPGKGTGTHGLGPKPKDCCGPPPGSTGTPGAGGEGGANGPQGRPPGHVGP 60  
QY 61 PGQGPAGISAVGLKDDRGATGGERGLAGLPQGPQPPGPGPGYGMKATGPMGQOQIPG 120  
DB 61 PGQGPAGISAVGLKDDRGATGGERGLAGLPQGPQPPGPGPGYGMKATGPMGQOQIPG 120  
QY 121 IPGPPGPMGQPGKAGHCNPSCFCGAMPMEQOYPPMTKMGPF 163  
DB 121 IPGPPGPMGQPGKAGHCNPSCFCGAMPMEQOYPPMTKMGPF 163

RESULT 7  
US-10-154-678-58  
; Sequence 58, Application US/10154678

```
/ Publication No. US20030162186A1
/ GENERAL INFORMATION:
/ APPLICANT: Benjamin, Stephane
/ APPLICANT: Tanaka, Hiroaki
/ TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
/ FILE REFERENCE: 182.US1.REG
/ CURRENT APPLICATION NUMBER: US/10/154,678
/ PRIOR FILING DATE: 2002-10-15
/ PRIOR APPLICATION NUMBER: US 09/924,340
/ PRIOR FILING DATE: 2001-08-06
/ PRIOR APPLICATION NUMBER: US 60/305,456
/ PRIOR FILING DATE: 2001-07-13
/ PRIOR APPLICATION NUMBER: US 60/302,277
/ PRIOR FILING DATE: 2001-06-29
/ PRIOR APPLICATION NUMBER: US 60/298,698
/ PRIOR FILING DATE: 2001-06-15
/ PRIOR APPLICATION NUMBER: US 60/293,574
/ PRIOR FILING DATE: 2001-05-25
/ NUMBER OF SEQ ID NOS: 112
/ SOFTWARE: JPatent
/ SEQ ID NO 58
/ LENGTH: 163
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-154-678-58
```

```
Query Match 100.0%; Score 952; DB 4; Length 163;
Best Local Similarity 100.0%; Pred. No. 7,4e-54;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MCPPGFKGTGHPGLPGPKDCCGKPPGPGSTGRPGAEGEPGAMPGQRPFGHVPFGP 60
DB 1 MCPPGFKGTGHPGLPGPKDCCGKPPGPGSTGRPGAEGEPGAMPGQRPFGHVPFGP 60
QY 61 PGQPGAGISAVGLKGDRCATGSRGLAGLPQGPFGPGPGPGYGMMGATGPMGQGGIFG 120
DB 61 PGQPGAGISAVGLKGDRCATGSRGLAGLPQGPFGPGPGPGYGMMGATGPMGQGGIFG 120
QY 121 IGGPPGMQPGKAGHCNPSDCFGAMPMEQYPPMKTMKGPF 163
DB 121 IGGPPGMQPGKAGHCNPSDCFGAMPMEQYPPMKTMKGPF 163
```

```
RESULT 8
US-10-838-854-58
/ Sequence 58, Application US/1083854
/ Publication No. US20050026182A1
/ GENERAL INFORMATION:
/ APPLICANT: Benjamin, Stephane
/ APPLICANT: Tanaka, Hiroaki
/ TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
/ FILE REFERENCE: 91.US6.DIV
/ CURRENT APPLICATION NUMBER: US/10/838,854
/ CURRENT FILING DATE: 2004-05-03
/ PRIOR APPLICATION NUMBER: US/10/000,489
/ PRIOR FILING DATE: 2001-11-14
/ PRIOR APPLICATION NUMBER: US 09/924,340
/ PRIOR FILING DATE: 2001-08-06
/ PRIOR APPLICATION NUMBER: PCT/IB01/01715
/ PRIOR FILING DATE: 2001-08-06
/ PRIOR APPLICATION NUMBER: US 60/305,456
/ PRIOR FILING DATE: 2001-07-13
/ PRIOR APPLICATION NUMBER: US 60/302,277
/ PRIOR FILING DATE: 2001-06-29
/ PRIOR APPLICATION NUMBER: US 60/298,698
/ PRIOR FILING DATE: 2001-06-15
/ PRIOR APPLICATION NUMBER: US 60/293,574
/ PRIOR FILING DATE: 2001-05-25
/ NUMBER OF SEQ ID NOS: 112
/ SOFTWARE: JPatent
/ SEQ ID NO 58
/ LENGTH: 163
/ TYPE: PRT
```

```
/ ORGANISM: Homo sapiens
US-10-838-854-58
```

```
Query Match 100.0%; Score 952; DB 5; Length 163;
Best Local Similarity 100.0%; Pred. No. 7,4e-54;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MCPPGFKGTGHPGLPGPKDCCGKPPGPGSTGRPGAEGEPGAMPGQRPFGHVPFGP 60
DB 1 MCPPGFKGTGHPGLPGPKDCCGKPPGPGSTGRPGAEGEPGAMPGQRPFGHVPFGP 60
QY 61 PGQPGAGISAVGLKGDRCATGSRGLAGLPQGPFGPGPGPGYGMMGATGPMGQGGIFG 120
DB 61 PGQPGAGISAVGLKGDRCATGSRGLAGLPQGPFGPGPGPGYGMMGATGPMGQGGIFG 120
QY 121 IGGPPGMQPGKAGHCNPSDCFGAMPMEQYPPMKTMKGPF 163
DB 121 IGGPPGMQPGKAGHCNPSDCFGAMPMEQYPPMKTMKGPF 163
```

```
RESULT 9
US-10-723-860-4225
```

```
/ Sequence 4225, Application US/10723860
/ Publication No. US20040253606A1
/ GENERAL INFORMATION:
```

```
/ APPLICANT: Aziz, Natasha
/ APPLICANT: Ginsburg, Wendy M.
```

```
/ APPLICANT: Zlotnik, Albert
```

```
/ TITLE OF INVENTION: Methods of diagnosis of Soft Tissue Sarcoma, Compositions &
```

```
/ TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
```

```
/ FILE REFERENCE: 05882.0193.NPUS01
```

```
/ CURRENT APPLICATION NUMBER: US/10/723,860
```

```
/ PRIOR FILING DATE: 2003-11-26
```

```
/ PRIOR APPLICATION NUMBER: 60/429,739
```

```
/ PRIOR FILING DATE: 2002-11-26
```

```
/ NUMBER OF SEQ ID NOS: 8393
```

```
/ SOFTWARE: PatentIn version 3.2
```

```
/ SEQ ID NO 4225
```

```
/ LENGTH: 1603
```

```
/ TYPE: PRT
```

```
/ ORGANISM: Homo sapiens
```

```
US-10-723-860-4225
```

```
Query Match 88.3%; Score 840.5; DB 5; Length 1603;
Best Local Similarity 44.5%; Pred. No. 6,5e-46;
Matches 163; Conservative 0; Mismatches 0; Indels 203; Gaps 1;
```

```
QY 1 MCPPGFKGTGHPGLPGPKDCCGKPPGPGSTGRPGAEGEPGAMPGQRPFGHVPFGP 60
DB 1238 MCPPGFKGTGHPGLPGPKDCCGKPPGPGSTGRPGAEGEPGAMPGQRPFGHVPFGP 1297
QY 61 PGQPGAGISAVGLKGDRCATGSRGLAGLPQGPFGPGPGPGYGMMGATGPMGQGGIFG 92
DB 1298 PGQPGAGISAVGLKGDRCATGSRGLAGLPQGPFGPGPGPGYGMMGATGPMGQGGIFG 1357
QY 93 ----- 92
DB 1358 GFYPPGPKDDBAAGQKQAGEKRGAPGPGKSGMSVPVPPGAPAGERHGPAGPS 1417
QY 93 ----- 92
DB 1418 GSPGLPGVPSMGDMVNYDEIRKTRIOELIKMFDERMAYTTSRMQPFMEMAAAPGRPGP 1477
QY 93 ----- 92
DB 1478 GNDGAPRGAPGSPGLPGQIGREGQGLPVYRGLPDTGKEKGDIGIAGENGGLPGPG 1537
QY 98 PGQPGYGKMGATGPMGQGGIFGIPGPFGPMQPGKAGHCNPSDCFGAMPMEQYPPMKT 157
DB 1538 PGQPGYGKMGATGPMGQGGIFGIPGPFGPMQPGKAGHCNPSDCFGAMPMEQYPPMKT 1597
QY 158 MKGPF 163
```



Matches 80; Conservative 11; Mismatches 40; Indels 54; Gaps 4;

```
QY 2 GPPGKGTGTHG-----LPGKDDCGKPPGSTGARGAGEPGAMPQGRPPGHV 55
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 147 GPPGIPGMKGEGSITMSLPGPKGNPGYPGPGIQLPPTGIPGP---PGPPGLM 203
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 56 GPPGPGGPAGISAVGLKGDGATGERGLAQLP-----YGRMGATGPMGQGIPIGTPGPMGQ 90
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 204 GPPGPGIPLPGPKGNMGLNFQPKGKEGEGLOGPPPGQISEQKRPIDVERQKDDQLP 263
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 91 ---GPPPPPGPPG-----YGRMGATGPMGQGIPIGTPGPMGQ 130
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 264 GDRGPPPGPIRGPPPGPKGKEGEGEPKRGKPGKDGEMGQPGIPLGPDGYPGE 323
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 131 PGKAG 135
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 324 PGRDG 328
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

## RESULT 14

```
US-10-450-763-35198
; Sequence 35198, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; PRIOR FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 35198
; LENGTH: 1685
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1159)..(1203)
; OTHER INFORMATION: Synapsins proteins domain identified by eMATRIX, accession
; OTHER INFORMATION: number BL00415N, p-value=4.229e-10, raw score of 4.29
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (42)..(1456)
; OTHER INFORMATION: Collagen triple helix repeat (20 copies) domain identified by
; OTHER INFORMATION: Pfam, accession name Collagen, E-value=3.2e-176, Pfam score of
; OTHER INFORMATION: 598.8
US-10-450-763-35198
```

Query Match 41.0%; Score 390; DB 5; Length 1685;

Best Local Similarity 43.2%; Pred. No. 4.7e-17;

Matches 80; Conservative 11; Mismatches 40; Indels 54; Gaps 4;

```
QY 2 GPPGKGTGTHG-----LPGKDDCGKPPGSTGARGAGEPGAMPQGRPPGHV 55
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 147 GPPGIPGMKGEGSITMSLPGPKGNPGYPGPGIQLPPTGIPGP---PGPPGLM 203
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 56 GPPGPGGPAGISAVGLKGDGATGERGLAQLP-----YGRMGATGPMGQGIPIGTPGPMGQ 90
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 204 GPPGPGIPLPGPKGNMGLNFQPKGKEGEGLOGPPPGQISEQKRPIDVERQKDDQLP 263
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 91 ---GPPPPPGPPG-----YGRMGATGPMGQGIPIGTPGPMGQ 130
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 264 GDRGPPPGPIRGPPPGPKGKEGEGEPKRGKPGKDGEMGQPGIPLGPDGYPGE 323
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 131 PGKAG 135
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 324 PGRDG 328
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

## RESULT 15

```
US-10-450-763-45978
; Sequence 45978, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; PRIOR FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 45978
; LENGTH: 1693
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1167)..(1211)
; OTHER INFORMATION: Synapsins proteins domain identified by eMATRIX, accession
; OTHER INFORMATION: number BL00415N, p-value=4.229e-10, raw score of 4.29
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (50)..(1462)
; OTHER INFORMATION: Collagen triple helix repeat (20 copies) domain identified by
; OTHER INFORMATION: Pfam, accession name Collagen, E-value=7.1e-185, Pfam score of
; OTHER INFORMATION: 627.6
US-10-450-763-45978
```

Query Match 41.0%; Score 390; DB 5; Length 1693;

Best Local Similarity 43.2%; Pred. No. 4.8e-17;

Matches 80; Conservative 11; Mismatches 40; Indels 54; Gaps 4;

```
QY 2 GPPGKGTGTHG-----LPGKDDCGKPPGSTGARGAGEPGAMPQGRPPGHV 55
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 155 GPPGIPGMKGEGSITMSLPGPKGNPGYPGPGIQLPPTGIPGP---PGPPGLM 211
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 56 GPPGPGGPAGISAVGLKGDGATGERGLAQLP-----YGRMGATGPMGQGIPIGTPGPMGQ 271
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 212 GPPGPGIPLPGPKGNMGLNFQPKGKEGEGLOGPPPGQISEQKRPIDVERQKDDQLP 271
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 91 ---GPPPPPGPPG-----YGRMGATGPMGQGIPIGTPGPMGQ 130
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 272 GDRGPPPGPIRGPPPGPKGKEGEGEPKRGKPGKDGEMGQPGIPLGPDGYPGE 331
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 131 PGKAG 135
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 332 PGRDG 336
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Search completed: December 1, 2005, 08:40:59

Job time : 163 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2005, 08:26:53 ; Search time 11 Seconds  
(Without alignments)  
70.954 Million cell updates/sec

Title: US-09-924-340-58  
Perfect score: 952  
Sequence: 1 MGPFGKGTGHTGHPGPKG.....GAMPMEQGYPMKTKMGPGF 163

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 26661 seqs, 4788334 residues  
Total number of hits satisfying chosen parameters: 26661

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : Published Applications\_AA\_New:  
1: /cgn2\_6/prodata/2/pubppa/US09\_NEW\_PUB.pcp:\*  
2: /cgn2\_6/prodata/2/pubppa/US06\_NEW\_PUB.pcp:\*  
3: /cgn2\_6/prodata/2/pubppa/US07\_NEW\_PUB.pcp:\*  
4: /cgn2\_6/prodata/2/pubppa/US08\_NEW\_PUB.pcp:\*  
5: /cgn2\_6/prodata/2/pubppa/PCT\_NEW\_PUB.pcp:\*  
6: /cgn2\_6/prodata/2/pubppa/US11\_NEW\_PUB.pcp:\*  
7: /cgn2\_6/prodata/2/pubppa/US11\_NEW\_PUB.pcp:\*  
8: /cgn2\_6/prodata/2/pubppa/US60\_NEW\_PUB.pcp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	385	40.4	1532	6	US-10-821-234-914	Sequence 914, App
2	383.5	40.3	1874	6	US-10-821-234-1182	Sequence 1182, App
3	377	39.6	1166	6	US-10-821-234-964	Sequence 964, App
4	372	39.1	1467	6	US-10-821-234-1096	Sequence 1096, App
5	356.5	37.4	1366	6	US-10-821-234-1431	Sequence 1431, App
6	245	25.7	303	7	US-11-135-855-31	Sequence 31, App1
7	239.5	25.2	406	6	US-10-131-826A-82	Sequence 82, App1
8	238.5	25.1	288	7	US-11-135-855-30	Sequence 30, App1
9	219.5	23.1	253	6	US-10-821-234-1438	Sequence 1438, App
10	209	22.0	636	6	US-10-793-626-2922	Sequence 2922, App
11	198	20.8	243	6	US-10-131-826A-362	Sequence 362, App
12	165.5	17.4	246	7	US-11-067-121-4	Sequence 4, App1
13	162.5	17.0	119	7	US-11-110-424-4	Sequence 4, App1
14	161.5	17.0	246	7	US-11-067-121-13	Sequence 13, App1
15	161.5	17.0	287	7	US-11-174-150-33	Sequence 33, App1
16	161	16.9	334	6	US-10-514-057-6	Sequence 6, App1
17	157	16.5	724	7	US-11-053-100-41	Sequence 41, App1
18	157	16.5	1174	7	US-11-053-100-43	Sequence 43, App1
19	152.5	16.0	669	7	US-11-053-100-39	Sequence 39, App1
20	149.5	15.7	489	7	US-11-053-100-24	Sequence 24, App1
21	149.5	15.7	504	7	US-11-053-100-28	Sequence 28, App1
22	149.5	15.7	506	7	US-11-053-100-37	Sequence 37, App1
23	149.5	15.7	507	7	US-11-053-100-31	Sequence 31, App1
24	149.5	15.7	507	7	US-11-053-100-34	Sequence 34, App1
25	149.5	15.7	939	7	US-11-053-100-26	Sequence 26, App1

26	149	15.7	879	7	US-11-053-100-55	Sequence 55, App1
27	149	15.7	1225	7	US-11-053-100-50	Sequence 50, App1
28	149	15.7	1329	7	US-11-053-100-56	Sequence 56, App1
29	145	15.2	864	7	US-11-053-100-58	Sequence 58, App1
30	144.5	15.2	624	7	US-11-053-100-48	Sequence 48, App1
31	144.5	15.2	641	6	US-10-848-976-1	Sequence 1, App1
32	144.5	15.2	728	7	US-11-053-100-54	Sequence 54, App1
33	144.5	15.2	735	7	US-11-053-100-45	Sequence 45, App1
34	144.5	15.2	736	7	US-11-053-100-46	Sequence 46, App1
35	144.5	15.2	757	7	US-11-053-100-47	Sequence 47, App1
36	144.5	15.2	774	7	US-11-053-100-49	Sequence 49, App1
37	144.5	15.2	775	7	US-11-053-100-51	Sequence 51, App1
38	144.5	15.2	859	7	US-11-053-100-52	Sequence 52, App1
39	144.5	15.2	879	7	US-11-053-100-57	Sequence 57, App1
40	144.5	15.2	1186	7	US-11-053-100-46	Sequence 46, App1
41	144.5	15.2	1309	7	US-11-053-100-53	Sequence 53, App1
42	144	15.1	288	6	US-10-131-826A-316	Sequence 316, App
43	141.5	14.9	66	6	US-10-514-057-1	Sequence 1, App1
44	141.5	14.9	284	7	US-11-102-240-62	Sequence 62, App1
45	141	14.8	334	6	US-10-802-796-728	Sequence 728, App
46	140.5	14.8	574	7	US-11-053-100-40	Sequence 40, App1
47	140.5	14.8	874	7	US-11-053-100-42	Sequence 42, App1
48	136.5	14.3	450	7	US-11-053-100-16	Sequence 16, App1
49	134	14.1	318	6	US-10-802-796-727	Sequence 727, App
50	133	14.0	250	6	US-10-821-234-1237	Sequence 1297, App
51	133	14.0	2630	7	US-11-186-731-2	Sequence 2, App1
52	133	14.0	7968	7	US-11-186-731-5	Sequence 36, App1
53	129.5	13.6	356	7	US-11-053-100-36	Sequence 36, App1
54	129.5	13.6	357	7	US-11-053-100-38	Sequence 38, App1
55	129.5	13.6	656	7	US-11-053-100-38	Sequence 38, App1
56	129.5	13.6	657	7	US-11-053-100-35	Sequence 35, App1
57	127.5	13.4	339	7	US-11-053-100-23	Sequence 23, App1
58	127.5	13.4	354	7	US-11-053-100-27	Sequence 27, App1
59	127.5	13.4	357	7	US-11-053-100-30	Sequence 30, App1
60	127.5	13.4	639	7	US-11-053-100-25	Sequence 25, App1
61	127.5	13.4	654	7	US-11-053-100-29	Sequence 29, App1
62	127.5	13.4	657	7	US-11-053-100-32	Sequence 32, App1
63	126.5	13.3	175	6	US-10-821-234-1074	Sequence 1074, App
64	122	12.8	278	7	US-11-113-424-43	Sequence 43, App1
65	120	12.6	1187	6	US-10-821-234-955	Sequence 955, App
66	119.5	12.6	100	7	US-11-053-100-17	Sequence 17, App1
67	117.5	12.3	440	7	US-11-102-240-52	Sequence 52, App1
68	117	12.3	298	6	US-10-821-234-899	Sequence 899, App
69	117	12.3	2432	6	US-10-821-234-1240	Sequence 1240, App
70	116	12.2	240	7	US-11-021-441-28	Sequence 28, App1
71	116	12.2	1377	6	US-10-821-234-1070	Sequence 1070, App
72	115.5	12.1	963	6	US-10-467-962B-2	Sequence 2, App1
73	114.5	12.0	614	7	US-11-015-546A-20	Sequence 20, App1
74	113	11.9	483	6	US-10-451-375-12	Sequence 12, App1
75	111.5	11.7	281	6	US-10-821-234-1032	Sequence 1032, App
76	109.5	11.5	552	6	US-10-821-234-1032	Sequence 1032, App
77	108	11.3	618	6	US-10-821-234-1481	Sequence 1481, App
78	105	11.0	158	6	US-10-467-962B-83	Sequence 83, App1
79	104.5	11.0	1560	7	US-11-059-982-1	Sequence 1, App1
80	103	10.8	431	7	US-11-058-735-79	Sequence 79, App1
81	102.5	10.8	1133	6	US-10-821-234-1219	Sequence 1219, App
82	100	10.5	199	7	US-11-113-424-44	Sequence 44, App1
83	99.5	10.5	401	6	US-10-821-234-881	Sequence 881, App
84	99.5	10.5	1970	6	US-10-821-234-1641	Sequence 1641, App
85	99	10.4	417	6	US-10-971-560-2	Sequence 2, App1
86	99	10.4	501	6	US-10-971-560-2	Sequence 2, App1
87	99	10.4	501	6	US-10-971-560-10	Sequence 10, App1
88	99	10.4	501	7	US-11-067-121-1	Sequence 1, App1
89	99	10.4	1618	6	US-10-984-645-2	Sequence 2, App1
90	98	10.3	273	6	US-10-821-234-1203	Sequence 1203, App
91	97	10.2	572	6	US-10-821-234-1250	Sequence 1250, App
92	95.5	10.0	190	6	US-10-131-826A-158	Sequence 158, App
93	95.5	10.0	283	6	US-10-467-657-3336	Sequence 3336, App
94	95.5	10.0	575	6	US-10-131-826A-128	Sequence 128, App
95	95	10.0	405	6	US-10-821-234-931	Sequence 931, App
96	94.5	9.9	341	6	US-10-821-234-1628	Sequence 1628, App
97	94.5	9.9	406	6	US-10-821-234-1026	Sequence 1026, App
98	94.5	9.9	494	6	US-10-971-560-7	Sequence 7, App1



```
Publication No. US20050255114A1
GENERAL INFORMATION:
APPLICANT: Labat, Ivan
APPLICANT: Stache-Crain, Birgit
APPLICANT: Andarmant, Susan
APPLICANT: Tang, Y. Tom
TITLE OF INVENTION: Methode for Diagnosis and Treatment of Preeclampsia
FILE REFERENCE: 821A
CURRENT APPLICATION NUMBER: US/10/821,234
PRIOR FILING DATE: 2004-04-07
PRIOR APPLICATION NUMBER: US 60/462,047
NUMBER OF SEQ ID NOS: 1704
SOFTWARE: pc_seq_genes Version 1.0
SEQ ID NO 1096
LENGTH: 1467
TYPE: PRT
ORGANISM: Homo sapiens
US-10-821-234-1096

Query Match
Best Local Similarity 39.1%; Score 372; DB 6; Length 1467;
Matches 81; Conservative 7; Mismatches 43; Indels 68; Gaps 6;

QY 2 GPP---GFKGKTGHPGLPGPKGDCGKPPGSTGTPGAESEPGAMGPGGHPGVPP 58
DB 872 GPGGATGFPAAARVGPSPGSGNAGPPGPPGPAKKEGKGPRTGTGAGRGVGP 931
QY 59 GP-----PGQPPGAGIS---AVGLKDRGATGBRGLAIP----- 90
DB 932 GPAGEKSPGADGPAAGTGPFGIAGRGVGLPQOR---GERGPPGLPGSPGEGKQ 988
QY 91 -----GQPPRPGQPPRGY-----GKMGATGPPGQQ 116
DB 989 GPGSASGERPPGPPGPPGLAGPPBSGREGAPAAEGSPGRDPSGAKGDRGSTGPP 1048
QY 117 GIPGIPGPPGPPMGQPPKAG 135
DB 1049 GARGARGAPGPPVGPAGKSG 1067

RESULT 5
US-10-821-234-1431
Sequence 1431, Application US/10821234
Publication No. US20050255114A1
GENERAL INFORMATION:
APPLICANT: Labat, Ivan
APPLICANT: Stache-Crain, Birgit
APPLICANT: Andarmant, Susan
APPLICANT: Tang, Y. Tom
TITLE OF INVENTION: Methode for Diagnosis and Treatment of Preeclampsia
FILE REFERENCE: 821A
CURRENT APPLICATION NUMBER: US/10/821,234
PRIOR FILING DATE: 2004-04-07
PRIOR APPLICATION NUMBER: US 60/462,047
PRIOR FILING DATE: 2003-04-07
NUMBER OF SEQ ID NOS: 1704
SOFTWARE: pc_seq_genes Version 1.0
SEQ ID NO 1431
LENGTH: 1366
TYPE: PRT
ORGANISM: Homo sapiens
US-10-821-234-1431

Query Match
Best Local Similarity 37.4%; Score 356.5; DB 6; Length 1366;
Matches 80; Conservative 16; Mismatches 67; Indels 25; Gaps 6;

QY 1 MGPPGKGTGHPGLPGPKDCKGPPRGSTGTPGAESEPGAMGPPGQPPRPP----- 51
DB 417 MGPPGKGTGHPGLPGPKDCKGPPRGSTGTPGAESEPGAMGPPGQPPRPP----- 51
QY 52 PGHVGPPGPPGPGGPPAGI-SAVGLKDRGATGBRGLAIPGQPPGPP-----GQPPG- 103
```

```
DB 477 PGPIDGARGEBGNIGFPQPKPTGDPKNGDKHAGLAGARGADPPGNNAGQPPG 536
QY 104 ---YGMKGATGPMGQGGIGIIPGPPMGQPPGAGHCNSTDCCG---AMPMPQYPPMK 156
DB 537 QGVQGGKGGQPPGPPGGLPSPGABGVKPPGRLHGERGLDPPAGPPRGPPG 596
QY 157 T-MKGPFG 163
DB 597 SGAAGPTG 604

RESULT 6
US-11-135-855-31
Sequence 31, Application US/1135855
Publication No. US2005025557A1
GENERAL INFORMATION:
APPLICANT: SMITHKLINE BEECHAM CORPORATION
APPLICANT: SMITHKLINE BEECHAM P.1.C.
TITLE OF INVENTION: NOVEL COMPOUNDS
FILE REFERENCE: GP50013
CURRENT APPLICATION NUMBER: US/11/135,855
PRIOR FILING DATE: 2005-05-24
PRIOR APPLICATION NUMBER: US/10/203,708
PRIOR FILING DATE: 2002-08-13
PRIOR APPLICATION NUMBER: PCT/US01/04703
PRIOR FILING DATE: 2001-02-14
PRIOR APPLICATION NUMBER: 60/182,172
PRIOR FILING DATE: 2000-02-14
PRIOR APPLICATION NUMBER: 60/186,084
PRIOR FILING DATE: 2000-02-29
NUMBER OF SEQ ID NOS: 46
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 31
LENGTH: 303
TYPE: PRT
ORGANISM: Homo sapiens
US-11-135-855-31

Query Match
Best Local Similarity 25.7%; Score 245; DB 7; Length 303;
Matches 53; Conservative 10; Mismatches 35; Indels 16; Gaps 4;

QY 13 PGLRPPGDCGKGGPPGSTGTPGAESE---PGAMGPPGPPGPPGPPGPPGPPGPPG 69
DB 51 PGLRPP-----PGPPGANGSPGPPHRTGLPGKDGDKKGGKGGKGTAGLRKTPGL 104
QY 70 SAVGLKDRGATGBRGLAIPGQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 123
DB 105 A--GKGDQGGTGGKPPIDGEGKGEVGPIDGP-----GPKGDRGEGDPPGLPG 151

RESULT 7
US-10-131-826A-82
Sequence 82, Application US/10131826A
Publication No. US20050245730A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerltzen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
```

APPLICANT Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME  
FILE REFERENCE: P3330R1C128  
CURRENT FILING DATE: 2002-04-24  
PRIOR APPLICATION NUMBER: US/10/131,826A  
PRIOR FILING DATE: 1997-06-18  
PRIOR APPLICATION NUMBER: 60/056974  
PRIOR FILING DATE: 1997-08-26  
PRIOR APPLICATION NUMBER: 60/059113  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059115  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059117  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059122  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059184  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/059263  
PRIOR FILING DATE: 1997-09-18  
PRIOR APPLICATION NUMBER: 60/059352  
PRIOR FILING DATE: 1997-09-19  
PRIOR APPLICATION NUMBER: 60/059588  
PRIOR FILING DATE: 1997-09-19  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 550

Query Match	25.2%	Score 239.5	DB 6	Length 406
-------------	-------	-------------	------	------------

Matches	52; Conservative	3; Mismatches	24; Indels	25; Gaps	4
Qy	2	GPFPKGTGTHPGI.PGPKGDCGKPPKPSTASTRPGAGEGPGAMGP-----	QGRPPGPG	53	
Db	247	GPPLPGGQGPSPGSGPFGKSGPCGFPMRPMRPPQGPGRGSMGMPGPPDLSHTKQGRGP--	304		
Qy	54	HVGPFGPPQGPQGPAGISAVGLKIDRRATGGRQLAGLPGQPPPG	97		
Db	305	-VGPPGAPGRDGS-----KGEKGAGGPR-----GSPGPG	333		

```

RESULT 8
US-11-135-855-30
; Sequence 30. Application US/11135855
; Publication No. US20050255557A1
GENERAL INFORMATION:
APPLICANT: SMITHKLINE BEECHAM CORPORATION
APPLICANT: SMITHKLINE BEECHAM P.L.C.
TITLE OF INVENTION: NOVEL COMPOUNDS
FILE REFERENCE: GP50013
CURRENT APPLICATION NUMBER: US/11/135,855
CURRENT FILING DATE: 2005-05-24
PRIORITY FILING DATE: US/10/203,708
PRIORITY FILING DATE: 2002-08-13
PRIORITY APPLICATION NUMBER: PCT/US01/04703
PRIORITY FILING DATE: 2001-02-14
PRIORITY APPLICATION NUMBER: 60/182,172
PRIORITY FILING DATE: 2000-02-14
PRIORITY APPLICATION NUMBER: 60/186,084
PRIORITY FILING DATE: 2000-02-29
NUMBER OF SEQ ID NOS: 46
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 30
;
; LENGTH: 288
;
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-11-135-855-30

```

Query Match	25.1%	Score 238.5	DB 7	Length 288
Best Local Similarity	46.0%	Pred. No. 8.9e-13		
Matches 52, Conservative 10, Mismatches 36				Indels 15; Gaps 4

QY 13 PGLPGKGDGCKGKPGKSTGSRPCAESEPPAMGQPGCPGVHGPSPG--GQPGAGIS 70  
 Db 37 PGLRGP-----PGPPGANGSPGPHRIQLPPIHGDGDKGKGEKGTALRGKTGPEGLA 90  
 QY 71 AVGLKDKDRATERGLAGLPQGPQPPGPGPGPGPYGSMATGTMGQGGIGTIG 123  
 Db 91 --GEKQGGTGGKGLPPEEGGKGVGP-----GPKDGRGGSDPGLRGL 136

```

RESULT 9
US-10-821-234-1438
; Sequence 1438, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Blyric
; APPLICANT: Andarmant, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 67/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ. ID NOS: 1704
; SOFTWARE: pc_seq_genes Version 1.0
; SEQ. ID NO 1438
; LENGTH: 253
; TYPE: PRF
; ORGANISM: Homo sapiens
; US-10-821-234-1438

```

Query Match	23.1%	Score 219.5	DB 6	Length 253
Best Local Similarity	44.6%	Pred. No. 2.4e-11		
Matches	45	Conservative 8	Indels 15	Gaps 2
QY	36	AEGERGAMGPGRRGPPGHVCPGPPGCGPAGISA----	VLGKDRGATGSEGLGAPG	91
	:::			
Db	25	SQAQLSCTGPPALPGIRPGIRPCTPPDQCPGPPGKIGEXGLDGLGHDGHEFGKGDPIPG		84
QY	92	QPGPGPGQPPGPGYKMGATGPMGGQGI	PGIRGPPGPMGQPG	132
Db	85	NPGKVGPK-----	GPWGKGGGAPGAPGPPKCESG	114

```

RESULT 10
US-10-793-626-2922
: Sequence 2922, Application US/10793626
: Publication No. US20050255478A1
: GENERAL INFORMATION:
: APPLICANT: KIMMELLY, WILLIAM JOHN
: TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
: FILE REFERENCE: P034800S
: CURRENT APPLICATION NUMBER: US/10/793,626
: CURRENT FILING DATE: 2004-03-04
: PRIOR APPLICATION NUMBER: 60/164,258
: PRIOR FILING DATE: 1999-11-09
: NUMBER OF SEQ ID NOS: 4472
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 2922
:
: LENGTH: 696
:
: TYPE: PRT
:
: ORGANISM: Artificial Sequence
:
: FEATURE:
:
: OTHER INFORMATION: Description of Artificial Sequence: synthetic
:
: US-10-793-626-2922

```

```

RESULT 10
US-10-793-626-2922
: Sequence 2922, Application US/10793626
: Publication No. US20050255478A1
: GENERAL INFORMATION:
: APPLICANT: KIMBERLY, WILLIAM JOHN
: TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
: FILE REFERENCE: PU348005
: CURRENT APPLICATION NUMBER: US/10/793,626
: CURRENT FILING DATE: 2004-03-04
: PRIOR APPLICATION NUMBER: 60/164,258
: PRIOR FILING DATE: 1999-11-09
: NUMBER OF SEQ ID NOS: 4472
: SOFTWARE: Patentin Ver. 2.1
: SEQ ID NO 2922
: LENGTH: 696
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: synthetic
: OTHER INFORMATION: amino acid sequence
US-10-793-626-2922

```

Query Match 22.0%; Score 209; DB 6; Length 696;  
Best Local Similarity 35.6%; Pred. No. 3.6e-10;  
Matches 48; Conservative 10; Mismatches 63; Indels 14; Gaps 4;

QY 2 GPPGFGKTKGHPGLPQPKDCCGKPPGPGSTGRGACGEPGAMPGQRPQPGHVGPPPP 61  
DB 520 GPP--PAEPKPKPAPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPP 577  
QY 62 GQPG-PAGISAVGLKQDRGATGGRGLAGLPQGPQPGQPGPGYGMKGATGPMGQGGIPG 120  
DB 578 AEGGTA-----EPGKPAEPKPAEPKPAEPKPAEPKPAEP--GKPAEPKPAEPKPA 626  
QY 121 IPGPPGPMGQPGKAG 135  
DB 627 EPKPPVPEPPTPAQSG 641

## RESULT 11

US-10-131-826A-362  
; Sequence 362, Application US/10131826A  
; Publication No. US20050245730A1

## GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.  
APPLICANT: Beresini, Maureen  
APPLICANT: DeForge, Laura  
APPLICANT: Desnoyers, Luc  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerltzen, Mary E.  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Sherwood, Steven  
APPLICANT: Smith, Victoria  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tuma, Daniel  
APPLICANT: Watanabe, Colin K  
APPLICANT: Wood, William  
APPLICANT: Zhang, Zemin

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE OF INVENTION: ACTDS ENCODING THE SAME

CURRENT APPLICATION NUMBER: US/10131, 826A

PRIOR FILING DATE: 2002-04-24

PRIOR APPLICATION NUMBER: 60/049911

PRIOR FILING DATE: 1997-06-18

PRIOR APPLICATION NUMBER: 60/056974

PRIOR FILING DATE: 1997-08-26

PRIOR APPLICATION NUMBER: 60/059113

PRIOR FILING DATE: 1997-09-17

PRIOR APPLICATION NUMBER: 60/059115

PRIOR FILING DATE: 1997-09-17

PRIOR APPLICATION NUMBER: 60/059117

PRIOR FILING DATE: 1997-09-17

PRIOR APPLICATION NUMBER: 60/059122

PRIOR FILING DATE: 1997-09-17

PRIOR APPLICATION NUMBER: 60/059184

PRIOR FILING DATE: 1997-09-17

PRIOR APPLICATION NUMBER: 60/059263

PRIOR FILING DATE: 1997-09-18

PRIOR APPLICATION NUMBER: 60/059352

PRIOR FILING DATE: 1997-09-19

PRIOR APPLICATION NUMBER: 60/059588

PRIOR FILING DATE: 1997-09-19

Remaining prior application data removed - See file wrapper or PALM.

NUMBER OF SEQ ID NOS: 550

SEQ ID NO 362

LENGTH: 243

TYPE: PRT

ORGANISM: Homo Sapien

US-10-131-826A-362

Query Match 20.8%; Score 198; DB 6; Length 243;

Best Local Similarity 32.2%; Pred. No. 1.1e-09;  
Matches 49; Conservative 6; Mismatches 41; Indels 56; Gaps 4;

QY 11 GHPGLGPPKGDCCGKPPGPGSTGRGACGEPGAMPGQRPQPGHVGPPPP 70  
DB 30 GHPGLGPPKGDCCGKPPGPGSTGRGACGEPGAMPGQRPQPGHVGPPPP 83  
QY 71 AVGLKQDRGATGGRGLAGLPQGPQPGQPGPGYGMKGATGPMGQGGIPG 130  
DB 84 -----GKPAEPKPAEPKPAEPKPAEPKPAEP--GKPAEPKPAEPKPA 101  
QY 131 PKAGHCNPSDFGAMPMEQYPPMKTKGPP 162  
DB 102 -----PRSAFSAKRSRVPSPDAPLPF 125

## RESULT 12

US-11-067-121-4  
; Sequence 4, Application US/11067121  
; Publication No. US20050261185A1

## GENERAL INFORMATION:

APPLICANT: Martijn, Cecile  
APPLICANT: Rondahl, Lena  
TITLE OF INVENTION: THERAPEUTIC PROTEINS  
FILE REFERENCE: 18909-002001  
CURRENT APPLICATION NUMBER: US/11/067, 121  
PRIOR FILING DATE: 2005-02-25  
PRIOR APPLICATION NUMBER: US 60/576, 445  
PRIOR FILING DATE: 2004-06-02  
PRIOR APPLICATION NUMBER: SE 0400489-1  
PRIOR FILING DATE: 2004-02-27  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: FaastSeq for Windows Version 4.0  
SEQ ID NO 4  
LENGTH: 246  
TYPE: PRT  
ORGANISM: Mus musculus

US-11-067-121-4

Query Match 17.4%; Score 165.5; DB 7; Length 246;

Best Local Similarity 38.5%; Pred. No. 3.5e-07;  
Matches 42; Conservative 7; Mismatches 31; Indels 29; Gaps 5;

QY 14 GLPGRKDCGKPPGPGSTGRGACGEPGAMPGQRPQPGHVGPPPPGAGISAVG 73  
DB 34 GLP--PDCSK-----CCHGDPGRGTPGPPGPP--GPPPPNHN----- 70  
QY 74 LKQDRGATGGRGLAGLPQGPQPGQPGPGYGMKGATGPMGQGGIPG 122  
DB 71 NGNNATGHBGAKGKDKGLGR-----GERGQHPKRGKGYGV 113

## RESULT 13

US-11-110-424-4  
; Sequence 4, Application US/11110424  
; Publication No. US20050261479A1

## GENERAL INFORMATION:

APPLICANT: Hoffmann, Christian K  
APPLICANT: Keller, Karsten  
TITLE OF INVENTION: A Method for Purifying and Recovering Silk Proteins Using  
FILE REFERENCE: CL2418 US NA  
CURRENT APPLICATION NUMBER: US/11/110, 424  
PRIOR FILING DATE: 2005-04-20  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: Patent version 3.2  
SEQ ID NO 4  
LENGTH: 119  
TYPE: PRT  
ORGANISM: Artificial Sequence

FEATURE:  
OTHER INFORMATION: Monomer of spider silk DP-2A analog protein

US-11-110-424-4

Query Match 17.1%; Score 162.5; DB 7; Length 119;  
Best Local Similarity 40.5%; Pred. No. 3.2e-07;  
Matches 53; Conservative 7; Mismatches 40; Indels 31; Gaps 10;

QY 4 PGFKKRTGH-PGLPPKRG--DCKRPP--PGSTGRGABSEPPAMGPPGPPGVHPGPG 59  
DB 15 PQQQGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 68  
QY 60 PPGQGPAGISAVGLKGRGATGERGLAGLPQGPQPPGPGPPGPGGPGGPGGPGGPGG 119  
DB 69 GPGQGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 110  
QY 120 GPGPPGPMGQ 130  
DB 111 --PGGYGP-GQ 118

## RESULT 14

US-11-067-121-13  
; Sequence 13, Application US/11067121  
; Publication No. US20050261185A1  
; GENERAL INFORMATION:  
; APPLICANT: Martijn, Cecile  
; APPLICANT: Rondahl, Lena  
; TITLE OF INVENTION: THERAPEUTIC PROTEINS  
; FILE REFERENCE: 18909-002001  
; CURRENT APPLICATION NUMBER: US/11/067,121  
; CURRENT FILING DATE: 2005-02-25  
; PRIOR APPLICATION NUMBER: US 60/576,445  
; PRIOR FILING DATE: 2004-06-02  
; PRIOR APPLICATION NUMBER: SE 0400489-1  
; PRIOR FILING DATE: 2004-02-27  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13  
; LENGTH: 246  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-067-121-13

Query Match 17.0%; Score 161.5; DB 7; Length 246;  
Best Local Similarity 38.6%; Pred. No. 7.1e-07;  
Matches 44; Conservative 7; Mismatches 32; Indels 31; Gaps 6;

QY 9 KTGHPGLPGPKGDCGKPPGPGSTGRGABSEPPAMGPPGPPGVHPGPGPGPGAG 68  
DB 31 QTC--GLP---PDCSK-----CCHGDYIFRGYQGPGRPP--GPPGIRGNHGN-- 70  
QY 69 ISAVGLKGRGATGERGLAGLPQGPQPPGPGPPGPGGPGGPGGPGGPGGPGGPGG 122  
DB 71 -----NGNNGATGHEGAKGEKGDGLGPR-----GERGQHGPKGEKGYPIG 113

## RESULT 15

US-11-174-150-30  
; Sequence 30, Application US/11174150  
; Publication No. US20050260714A1  
; GENERAL INFORMATION:  
; APPLICANT: Agarwal, Pankaj  
; APPLICANT: Murdoch, Paul R.  
; APPLICANT: Rizvi, Safia K.  
; APPLICANT: Smith, Randall F.  
; APPLICANT: Xiang, Zhaoying  
; TITLE OF INVENTION: NOVEL COMPOUNDS  
; FILE REFERENCE: GP50022  
; CURRENT APPLICATION NUMBER: US/11/174,150  
; CURRENT FILING DATE: 2005-07-01  
; PRIOR APPLICATION NUMBER: US/10/257,174  
; PRIOR FILING DATE: 2002-10-10  
; PRIOR APPLICATION NUMBER: PCT/US01/11797  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 60/196,603

; PRIOR FILING DATE: 2000-04-13  
; PRIOR APPLICATION NUMBER: 60/199,417  
; PRIOR FILING DATE: 2000-04-24  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 30  
; LENGTH: 287  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-174-150-30

Query Match 17.0%; Score 161.5; DB 7; Length 287;  
Best Local Similarity 36.8%; Pred. No. 8.1e-07;  
Matches 43; Conservative 10; Mismatches 29; Indels 35; Gaps 6;

QY 27 PPGST-----GRGABSEPPAMGPPGPPGVHPGPPGPGPGGPGGPGGPGGPGG 71  
DB 17 PPGAHYEMLTGTCRMICDPYSVAAPAG--GPAKAKAP-----PGSTALEVMODLSNP 69  
QY 72 -----VGLKGRGATGERGLAGLPQGPQPPGPGPPGPGGPGGPGGPGGPGGPGG 123  
DB 70 PPPFIQGPGRGDPGRPKGPPGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGGPGG 118

Search completed: December 1, 2005, 08:38:10  
Job time : 12 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2005, 08:18:41 ; Search time 16 Seconds  
(Without alignments)  
980.208 Million cell updates/sec

Title: US-09-924-340-58  
Perfect score: 952  
Sequence: 1 MGPPEFGKGTGHPGLDGPVK.....GAMPMEQGYPPMKTKGPFQ 163

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	840.5	88.3	1603	2 S23810	collagen alpha 1(X
2	390	41.0	1691	1 S22917	collagen alpha 5(I
3	387	40.7	1042	1 CGCH15	collagen alpha 1(I
4	387	40.7	1146	2 A38587	collagen, cornea-s
5	385	40.4	1532	2 A61262	collagen alpha 1(X
6	384	40.3	1549	2 I48103	type VII collagen
7	383	40.2	674	2 S13301	collagen alpha 1(X
8	382.5	40.2	1049	1 CGB075	collagen alpha 1(I
9	382.5	40.2	1838	1 CGH01V	collagen alpha 1(V
10	382.5	40.2	1843	2 S18803	collagen alpha 1(V
11	380.5	40.0	1433	2 A46053	bullous pemphigoid
12	379	39.8	488	2 A27353	collagen alpha 1(I
13	377.5	39.7	674	2 S23297	collagen alpha 1(X
14	377.5	39.7	888	2 S28791	collagen alpha 1(X
15	377.5	39.6	1466	1 CGH07L	collagen alpha 1(I
16	376.5	39.5	1690	1 CGH01B	collagen alpha 4(I
17	376	39.5	886	2 I50694	collagen alpha 1(I
18	374	39.3	921	2 S42617	collagen alpha 1(I
19	374	39.3	1464	2 S59856	collagen alpha 1(I
20	372.5	39.1	680	1 CGH01D	collagen alpha 1(X
21	372.5	39.1	684	2 A53019	collagen alpha 1(X
22	372	39.1	920	2 B34493	collagen alpha 1(I
23	371.5	39.0	730	2 A36226	collagen alpha 1(X
24	371.5	39.0	1806	1 CGH01B	collagen alpha 1(X
25	371	39.0	636	2 S41067	collagen alpha 1(I
26	371	39.0	779	1 CGB015	collagen alpha 1(I
27	370.5	38.9	1707	2 A33526	collagen alpha 2(I
28	370.5	38.9	1419	2 A41182	collagen alpha 1(I
29	370.5	38.9	1487	2 B41182	collagen alpha 1(I

30	370	38.9	921	2 S40495	collagen alpha 1(I
31	370	38.9	1453	2 S21626	collagen alpha 1(I
32	370	38.9	1464	1 CGH015	collagen alpha 1(I
33	369.5	38.8	1418	2 T45467	collagen alpha 1(I
34	369.5	38.8	1487	1 CGH06C	collagen alpha 1(I
35	368.5	38.7	2944	2 A54849	collagen alpha 1(V
36	368	38.7	1669	1 CGH04B	collagen alpha 1(I
37	367.5	38.6	1497	2 I49607	procollagen type V
38	367	38.6	1388	2 A53117	collagen alpha 1(X
39	366.5	38.5	615	2 A05269	collagen alpha 1(I
40	366.5	38.5	744	1 A34246	collagen alpha 1(V
41	366.5	38.5	744	1 S23298	collagen alpha 1(V
42	366	38.4	325	2 S02170	collagen alpha 1(I
43	366	38.4	473	2 I50629	collagen - chicken
44	365.5	38.4	453	2 S18804	collagen alpha 4(I
45	365	38.3	680	2 S32436	collagen alpha 2(I
46	365	38.3	680	2 S31216	collagen alpha 1(X
47	365	38.3	754	2 A55267	collagen alpha 5(I
48	364.5	38.3	469	2 A24450	collagen alpha 2(V
49	364.5	38.3	635	2 A57131	collagen alpha 2(V
50	364	38.2	198	2 I49558	collagen alpha 1(I
51	363.5	38.2	633	2 B40983	collagen alpha 1(X
52	363	38.1	920	2 A45748	collagen alpha 1(V
53	363	38.1	1496	1 CGH02V	collagen alpha 2(V
54	362.5	38.1	302	2 T15936	hypothetical prote
55	362.5	38.1	1546	1 CGH02E	collagen alpha 2(X
56	362	38.0	1763	2 S16366	collagen alpha 2(I
57	361.5	38.0	310	2 I50696	collagen alpha 1(I
58	361.5	38.0	677	2 S23296	collagen alpha 2(I
59	361.5	38.0	1024	2 S18251	collagen alpha 1(X
60	361.5	38.0	1142	2 JX0369	collagen alpha 1(X
61	361	37.9	482	2 B31795	collagen alpha 1(X
62	361	37.9	1373	1 A43291	collagen alpha 2(I
63	359.5	37.8	1669	1 CGH04B	collagen alpha 1(I
64	359.5	37.8	931	2 S13580	collagen alpha 1(I
65	359.5	37.8	1414	1 S23809	collagen alpha 2(I
66	359	37.7	671	1 CGH01S	collagen alpha 1(I
67	359	37.7	688	2 A53130	collagen alpha 2(I
68	358	37.6	3179	2 A31219	collagen 1 - Caeno
69	357.5	37.6	291	2 T20942	hypothetical prote
70	357.5	37.6	623	2 A45137	collagen alpha 4(I
71	357.5	37.6	743	1 S23779	collagen alpha 1(V
72	357.5	37.6	1670	1 CGH03B	collagen alpha 3(I
73	357	37.5	252	2 A55047	collagen alpha 1(V
74	357	37.5	301	2 T21314	hypothetical prote
75	357	37.5	675	2 S20819	collagen alpha 3(I
76	356.5	37.4	673	1 CGB06C	collagen alpha 1(I
77	356	37.4	1492	2 A40333	collagen alpha 1(I
78	355.5	37.3	1366	1 CGH02S	collagen alpha 2(I
79	355	37.3	290	2 A32249	collagen - sea urc
80	354.5	37.2	295	2 A44984	collagen - nematod
81	354.5	37.2	299	2 T29956	hypothetical prote
82	354	37.2	3124	2 A40020	collagen alpha 1(X
83	353	37.1	1486	1 B40333	collagen alpha 1(I
84	350.5	36.8	286	2 S34665	collagen, cuticula
85	350.5	36.8	964	1 CGH02S	collagen alpha 2(I
86	350	36.8	1759	2 T29351	collagen alpha 2(I
87	349	36.7	1315	2 A56101	hypothetical prote
88	349	36.7	1774	2 B56101	collagen alpha 1(X
89	348.5	36.6	318	2 S27977	cuticle collagen d
90	348.5	36.6	1758	2 T29350	hypothetical prote
91	348.5	36.6	1759	2 T29351	collagen alpha 2(I
92	348	36.6	1775	2 A31893	collagen alpha 1(I
93	346	36.3	744	2 S15435	collagen alpha 1(V
94	345.5	36.3	516	2 C44479	scavenger receptor
95	345	36.2	742	2 JC7595	collagen alpha 6(I
96	345	36.2	1691	1 CGH06B	hypothetical prote
97	344.5	36.2	298	2 T27644	collagen COL1 - f
98	344.5	36.2	812	2 S31521	collagen alpha 2(I
99	344	36.1	1712	1 CGH02B	collagen alpha 1 c
100	344	36.1	632	2 S42731	collagen alpha 1 c

## ALIGNMENTS

RESULT 1  
S23810  
collagen alpha 1(XVI) chain precursor - human  
N:Alternate names: procollagen alpha 1(XVI) chain  
C:Species: Homo sapiens (man)  
C:Date: 28-Oct-1994 #sequence, revision 28-Oct-1994 #text\_change 09-Jul-2004  
C:Accession: S23810, P00612, S08012  
R:Pan, T.C.; Zhang, R.Z.; Mattei, M.G.; Timpl, R.; Chu, M.L.  
Proc. Natl. Acad. Sci. U.S.A. 89, 6565-6569, 1992  
A:Title: Cloning and chromosomal location of human alpha1(XVI) collagen.  
A:Reference number: S23810; MUID:92335339; PMID:1631157  
A:Accession: S23810  
A:Molecule type: mRNA  
A:Residues: 1-1603 <PAN>  
A:Cross-references: UNIPROT:007092; UNIPARC:UPI0000126D33; EMBL:M92642; NID:g180757; PID:R:Experimental source: skin fibroblasts  
R:Yamaguchi, N.; Kimura, S.; McBride, O.W.; Hori, H.; Yamada, Y.; Kanamori, T.; Yamakoshi, J. Biochem. 112, 856-863, 1992  
A:Title: Molecular cloning and partial characterization of a novel collagen chain, alpha A:Reference number: P00612, MUID:93203161; PMID:1284248  
A:Accession: P00612  
A:Molecule type: mRNA  
A:Residues: 'GGR', 421-536, 'P', 538-1159, 'P', 1161-1162, 'P', 1164, 'P', 1166-1603 <YAN>  
A:Cross-references: UNIPARC:UPI000016B3C2; GB:S57132; NID:g298641; PIDN:AAB25797.1; PID:R:Experimental source: placenta  
R:Kimura, S.  
submitted to the EMBL Data Library, April 1989  
A:Description: Partial nucleotide and amino acid sequence of a collagen-like protein from A:Reference number: S08012  
A:Accession: S08012  
A:Molecule type: mRNA  
A:Residues: 403-419, 'GR', 421-536, 'P', 538-846, 'VM', <KIM>  
A:Cross-references: UNIPARC:UPI0000073DBB; EMBL:X14963; NID:g29984; PIDN:CAA33085.1; PID:R:Comment: Prolines and lysines at the third position of the tripeptide repeating unit edited and subsequently O-glycosylated.  
C:Genetics:  
A:Gene: GDB:COL16A1  
A:Cross-references: GDB:134045; OMIM:120326  
A:Map position: 1p34-1p34  
C:Complex: type XVI collagen may be a homotrimer, or a heterotrimer of two alpha 1(XVI) C:Function:  
A:Description: structural component of extracellular fibrous polymer as a minor form pro A:Note: may play a role in forming elastic connections at fibril surfaces  
C:Keywords: cell binding; coiled coil; extracellular matrix; glycoprotein; hydroxyllysine  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-1603/Product: collagen alpha 1(XVI) chain #status predicted <MAT>  
F:22-333/Region: amino-terminal nonhelical #status predicted <NC1>  
F:334-1577/Region: interrupted helical  
F:334-360/Domain: collagenous COL9 #status predicted <COL9>  
F:375-505/Domain: collagenous COL9 #status predicted <COL9>  
F:521-554/Domain: collagenous COL8 #status predicted <COL8>  
F:539-541/Region: cell attachment (R-G-D) motif  
F:572-630/Domain: collagenous COL7 #status predicted <COL7>  
F:652-722/Domain: collagenous COL6 #status predicted <COL6>  
F:738-875/Domain: collagenous COL5 #status predicted <COL5>  
F:887-938/Domain: collagenous COL4 #status predicted <COL4>  
F:973-987/Domain: collagenous COL3 #status predicted <COL3>  
F:1005-1007/Region: cell attachment (R-G-D) motif  
F:1011-1432/Domain: collagenous COL2 #status predicted <COL2>  
F:1226-1228/Region: cell attachment (R-G-D) motif  
F:1472-1577/Domain: collagenous COL1 #status predicted <COL1>  
F:1578-1603/Domain: carboxyl-terminal nonhelical #status predicted <NC01>  
F:7,327/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 88.3%; Score 840.5; DB 2; Length 1603;  
Best Local Similarity 44.5%; Pred. No. 4,4e-46;  
Matches 163; Conservative 0; Mismatches 0; Indels 203; Gaps 1;

QY 1 MGPPEKGTGHPGLPGPKDCKRPGPGSTGPGABGEPGAMGPGRPGPHVGPBP 60  
|||||

Db 1238 MGPPEKGTGHPGLPGPKDCKRPGPGSTGPGABGEPGAMGPGRPGPHVGPBP 1297  
QY 61 PGQPGAGISAVGLKGDRCATGERGLAGLPGQ----- 92  
Db 1298 PGQPGAGISAVGLKGDRCATGERGLAGLPGQPGPHGPPGPBPTDGAAGKGGPKQ 1357  
QY 93 ----- 92  
Db 1358 GFYPPGPKGDGAAGQKQAGEKGRAGMPGPGKSGMGPVPPGPPAGERGHPGAPGS 1417  
QY 93 ----- 92  
Db 1418 GSPGLPGVSGMDWVNDIEIKRPIRQELIKMFDERMAYYTSRMQPMEMAAAPGRGPP 1477  
QY 93 -----PGPPG 97  
Db 1478 GKDAAGPRGAPGSPGLPGQIGREGRGQLPVRGLPGTKGEXKDIDIGIAGENGGLPGPPG 1537  
QY 98 PGQPGPYGKMGATGPMGQGGIPGIPRPGPMGQPKAGHCNPSDRCGAMPMEQYPPMKT 157  
Db 1538 PGQPGPYGKMGATGPMGQGGIPGIPRPGPMGQPKAGHCNPSDRCGAMPMEQYPPMKT 1597  
QY 158 MKGPPG 163  
Db 1598 MKGPPG 1603

## RESULT 2

S22917  
collagen alpha 5(IV) chain precursor, renal splice form - human  
N:Alternate names: procollagen alpha 5(IV) chain  
N:Contains: collagen alpha 5(IV) chain precursor, leukocyte splice form  
C:Species: Homo sapiens (man)  
C:Date: 30-Sep-1993 #sequence, revision 27-Feb-1997 #text\_change 09-Jul-2004  
C:Accession: S22917, A54365; A57079; A37122; I54312; A34850; S18850; I56971; I76598; A35:R:Zhou, J.; Hertz, J.M.; Leinonen, A.; Trygsvaen, K.  
J. Biol. Chem. 267, 12475-12481, 1992  
A:Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and identical Alport syndrome patient.  
A:Reference number: S22917; MUID:92316923; PMID:1352287  
A:Accession: S22917  
A:Molecule type: mRNA  
A:Residues: 1-967 <ZHO>  
A:Cross-references: UNIPROT:P29400; UNIPARC:UPI0000173BDF; GB:M90464; NID:g180826; PIDN:R:Zhou, J.; Leinonen, A.; Trygsvaen, K.  
J. Biol. Chem. 269, 6608-6614, 1994  
A:Title: Structure of the human type IV collagen COL4A5 gene.  
A:Reference number: A54365; MUID:94165049; PMID:8120014  
A:Accession: A54365  
A:Molecule type: DNA  
A:Residues: 1-922 <ZHO>  
A:Cross-references: UNIPARC:UPI0000173BE0; GB:U04470; NID:g463378; GB:U04520; NID:g46342R:Zhou, J.; Mochizuki, T.; Smeets, H.; Antignac, C.; Laurila, P.; de Paeye, A.; Trygsvaen Science 261, 1167-1169, 1993  
A:Title: Deletion of the paired alphas5(IV) and alpha5(IV) collagen genes in inherited smc A:Reference number: A57079; MUID:93361972; PMID:8356449  
A:Accession: A57079  
A:Molecule type: DNA  
A:Residues: 1-27 <ZHO>  
A:Cross-references: UNIPARC:UPI000007378A; GB:Z37153; NID:g587203; PIDN:CA85512.1; PID:R:PhilaJantem, T.; Pohjolaime, E.R.; Myers, J.C.  
J. Biol. Chem. 265, 13758-13766, 1990  
A:Title: Complete primary structure of the triple-helical region and the carboxyl-terminal A:Reference number: A37122; MUID:90337990; PMID:2380186  
A:Accession: A37122  
A:Molecule type: mRNA  
A:Residues: 84-439, 'GS', 442-624, 'LAIQ', 629-666, 'RR', 669-887, 'R', 889-1264, 1271-1691 <PIH>  
A:Cross-references: UNIPARC:UPI0000173BE1; GB:J05558; EMBL:M58526; NID:g1314209  
A:Note: submitted to the EMBL Data Library, February 1991  
R:Remieri, A.; Seri, M.; Myers, J.C.; PhilaJantem, T.; Massella, L.; Rizzoni, G.; De Mai Hum. Mol. Genet. 1, 127-129, 1992  
A:Title: De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in f



A:Reference number: 154317; MUID:93244772; PMID:1363780  
A:Accession: 154317  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 313-324, 'E', 326-330 <REN>  
A:Cross-references: UNIPARC:UP1000016B3D0; GB:S59334; NID:g299946; PIDN:AAD13909.1; PID:R.Hochlika, S.L.; Eddy, R.L.; Byers, M.G.; Hoehlytae, M.; Shows, T.B.; Trygsvaen, K.  
Proc. Natl. Acad. Sci. U.S.A. 87, 1606-1610, 1990  
A>Title: Identification of a distinct type IV collagen alpha chain with restricted kidney  
A:Reference number: A34850; MUID:90160375; PMID:1689491  
A:Accession: A34850  
A:Molecule type: mRNA  
A:Residues: 914-1264, 1271-1691 <HOS>  
A:Cross-references: UNIPARC:UP1000016A70B; EMBL:M31115; NID:g180824; PIDN:AAA52045.1; PID:R.Zhou, J.; Hochlika, S.L.; Chow, L.T.; Trygsvaen, K.  
Genomics 9, 1-9, 1991  
A>Title: Characterization of the 3' half of the human type IV collagen alpha-5 gene chara  
A:Reference number: A37969; MUID:91169491; PMID:2004755  
A:Accession: 318850  
A:Molecule type: DNA  
A:Residues: 924-1264, 1271-1691 <ZH3>  
A:Cross-references: UNIPARC:UP1000016A437; EMBL:M63455; EMBL:M63456; EMBL:M63457; EMBL:M63458; EMBL:M63468; EMBL:M63470; EMBL:M63471; EMBL:M63472; EMBL:M63473; NID:g1779292  
R.Guo, C.; Van Damme, B.; Van Damme-Lombaerle, R.; Van den Berghe, H.; Cassiman, J.J.; M  
Kidney Int. 44, 1316-1321, 1993  
A>Title: Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex  
A:Reference number: 156971; MUID:94133540; PMID:8301933  
A:Accession: 156971  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1258-1276 <GUO1>  
A:Cross-references: UNIPARC:UP1000016B421; GB:S69168; NID:g545095; PIDN:AAC60612.1; PID:A:Accession: 176598  
A>Note: kidney splice form  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1284-1291, 'TFLGYLACLV', <GUO2>  
A:Cross-references: UNIPARC:UP1000011DDFD; GB:S69169; NID:g545097; PIDN:AAC60613.1; PID:A:Note: frameshift mutation in patient with Alport syndrome.  
R.Wyers, J.C.; Jones, T.A.; Pohjalainen, E.R.; Kadri, A.S.; Goddard, A.D.; Sheer, D.; Sc  
Am. J. Hum. Genet. 46, 1024-1033, 1990  
A>Title: Molecular cloning of alpha5(V) collagen and assignment of the gene to the regi  
A:Reference number: A35335; MUID:90252791; PMID:2339699  
A:Accession: A35335  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1448-1477 <MYE>  
A:Cross-references: UNIPARC:UP10000173BE2  
R.Nakazato, H.; Hattori, S.; Ushijima, T.; Matsura, T.; Kotabaehi, Y.; Takada, T.; Yoe  
Kidney Int. 46, 1307-1314, 1994  
A>Title: Mutations in the COL4A5 gene in Alport syndrome: a possible mutation in primord  
A:Reference number: 156975; MUID:95156893; PMID:7853788  
A:Accession: 156975  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1595-1602 <NAK>  
A:Cross-references: UNIPARC:UP10000004F8; GB:S75903; NID:g913882; PIDN:AA33374.1; PID:A:Note: premature termination mutation from a patient with Alport syndrome; one other mu  
R.Lemliak, H.H.; Schroeder, C.H.; Brunner, H.G.; Nelen, M.R.; Zhou, J.; Trygsvaen, K.;  
Genomics 17, 485-489, 1993  
A>Title: Identification of four novel mutations in the COL4A5 gene of patients with Alpo  
A:Reference number: 154188; MUID:94010948; PMID:8406498  
A:Accession: 154188  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1604-1607, 'VHDAKYC', <LEM>  
A:Cross-references: UNIPARC:UP100001185C; GB:S65767; NID:g425563; PIDN:AAD13967.1; PID:A:Note: frameshift mutation in a patient with Alport syndrome; five other mutations an  
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (e  
ed and subsequently O-glycosylated.  
C:Genetics:  
A:Gene: GDB:COL4A5; ATS  
A:Cross-references: GDB:120596; OMIM:303630

A:Map position:Xq22-XX722  
A:Introms: 27/3; 47/3; Xq22-8923; 92/3; 107/3; 128/3; 146/3; 155/3; 182/3; 203/3; 215/3; 229/3;  
3;/ 799/1; 837/1; 893/1; 923/1; 973/1; 1006/1; 1036/1; 1082/3; 1125/1; 1152/1; 1185/1; 121

A>Note: The alpha 5(IV) and alpha 6(IV) chain genes are encoded on opposite strands with  
C:Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 5(IV)  
er trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric am  
ter associations in the interrupted helical domain (with disulfide and desmosine cross-lin

C:Function:  
Description: minor structural component of extracellular basement membrane  
C:Superfamily: collagen alpha 1(IV) chain  
C:Keywords: Alport syndrome; basement membrane; coiled coil; extracellular matrix; glycop  
R:1-26/Domain: signal sequence #status predicted <SIG>  
F:27-1691/Product: collagen alpha 5(IV) chain, renal splice form #status predicted <MAT1>  
F:27-1264,1271-1691/Product: collagen alpha 5(IV) chain, leukocyte splice form #status p  
F:27-41/Domain: amino-terminal nonhelical, NC2 #status predicted <NC2>  
F:142-1462/Region: Interrupted helical  
F:1463-1691/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>  
F:1473-1573/Domain: collagen IV carboxyl-terminal repeat <CTR1>  
F:1583-1687/Domain: collagen IV carboxyl-terminal repeat <CTR2>  
F:129\_32\_38\_40\_124\_451\_481\_484/Dsulfide bonds: interchain #status predicted  
F:125/Binding site: carbonylcarate (asn) (covalent) #status predicted  
F:1482-1570,1515-1573/Dsulfide bonds: (or 1482-1573, 1515-1570) #status predicted  
F:1527-1533,1638-1644/Dsulfide bonds: #status predicted  
F:1592-1684,1626-1687/Disulfide bonds: (or 1592-1687, 1626-1684) #status predicted

Query Match            41.0%; Score 390; DB 1; Length 1691;  
Best Local Similarity     43.2%; Pred.No. 1,2e+17;  
Matches      80; Conservative    11; Mismatches    40; Indels       54; Gaps          4;

OY            2   GPFGKGTGTHPG-----LPPKKDDCGKRPGSGTGRPBAGEPQGMCPQRSPRGHV    55  
DB           147   GPGLPGMKKEBGGITIMSSLPPPKSNPYGPBPGLIGLPPTGIPIPIIP--PFPPLM    203  
  
OY            56   GPFGPGGPGAPAGISAVGLKGDRGATGERKLGLP-----YGKMGA TPWGQGIGIPIGPBPBMGO    130  
DB           204   GPFGPGLPGFKGNMGNLNPFQPKGEKESGGLGPPRPPIPQSIRKKRIDVRFQKGDGPL    263  
  
DB           264   GDRCPPGPPGRI RPPGPGEGEKERKDEQPPEKRRKPXKDENGNGGIPGLPDGDPGYGE    323  
  
OY            131   PKKAG    135  
DB           324   PFDG    328

RESULT 3  
CGCHS  
collagen alpha 1(I) chain - chicken (tentative sequence) (fragments)  
C:Species: Gallus gallus (chicken)  
C>Date: 12-Aug-1981 #sequence revision 06-Jul-1982 #text\_change 31-Mar-2000  
C:Accession: A90458; A90181; A02857  
R:Higberger, J.A.H.; Corbett, C.; Dixit, S.N.; Yu, W.; Seyler, J.M.; Kang, A.H.; Gross, J.  
Biochemistry 21, 2048-2055, 1982  
A>Title: Amino acid sequence of chick skin collagen alpa1(I)-CB8 and the complete prima  
A:Reference number: A90458; PMID:82231995; PMID:7093229  
A:Accession: A90458  
A:Molecule type: protein  
A:Residues: 1-1036 <HTG>  
A:CROSS-references: UNIPARC:UPI0000173B62  
A:Experimental source: skin  
A>Note: This is the latest in a series of papers from these workers elucidating the sequen  
R:Eyre, D.R.; Glimcher, M.J.  
Biochem. Biophys. Res. Commun. 48, 720-726, 1972  
A>Title: Evidence from a previously undetected sequence at the carboxyterminus of the alp  
A:Reference number: A90181; PMID:72243016; PMID:5047697  
A:Accession: A90181  
A:Molecule type: protein  
A:Residues: 1037-1042 <EVR>  
A:CROSS-references: UNIPARC:UPI0000173B63  
A:Experimental source: skin  
A>Note: Residues 1037-1042 above correspond to the carboxyl end of the protein  
A:Comment: Lysines at positions 103, 700, 934, and 946 above may be hydroxylated in some



Db 83 GPKGDRGCTGTGPEPEKGERGPPGVGPGGLPGVAGHPVEGPEGPPGPPGAGRGKGPB 142  
Qy 62 GGP-----GPAISAVGLKQDRGATGERGLAGLPQ-----PGPPGQGPFG----- 103  
Db 143 GRGDPVAVPGGAGAGKGERKQDAGLPGRKGAAGIKGQGPGLALPDPPGKDPGDRGPI 202  
Qy 104 -----YKMGATGPMQOGIPGIPGPPGPMQGPAGKHC-----NPSD 141  
Db 203 GLTRAGPFGDSCPPGKGDPPGRPPGPPVGSRRGRGVEKGEVGNPGD 252

RESULT 7  
S13301  
collagen alpha 1(X) chain precursor - bovine  
C:Species: Bos primigenius taurus (cattle)  
C>Date: 21-Nov-1993 #sequence\_revision 23-Feb-1996 #text\_change 09-Jul-2004  
C:Accession: S13301  
R:Thomas, J.T.; Kwan, A.P.L.; Grant, M.E.; Boot-Handford, R.P.  
Biochem. J. 273, 141-148, 1991  
A>Title: Isolation of cDNAs encoding the complete sequence of bovine type X collagen. EV  
A:Reference number: S13301; MUID:91131311; PMID:1703407  
A:Accession: S13301  
A:Molecule type: mRNA  
A:Residues: 1-674 <THO>  
A:Cross-references: UNIPROT:P23206; UNIPARC:UPI0000126D26; EMBL:X53556; NID:9263; PIDN:C  
C:Genetics:  
A:Gene: COL10A1  
C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology  
C:Keywords: coiled coil; extracellular matrix; glycoprotein; homotrimer  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-674/Product: collagen alpha 1(X) chain #status predicted <MAT>  
F:547-673/Domain: complement C1q carboxyl-terminal homology <CIQ>

Query Match 40.2%; Score 383; DB 2; Length 674;  
Best Local Similarity 35.0%; Pred. No. 1.6e-17;  
Matches 92; Conservative 8; Mismatches 51; Indels 112; Gaps 6;

Qy 2 GPPGFKGKTGHP-----GLGPKGDDCGKPPPG--STGRPGAGEPPGAMPOGR 48  
Db 63 GPPGPRGPRGHGPPSPGPPKPGSGSPGQGPGLPFPGPSATGKGLGPKGGERGL 122  
Qy 49 GPP-----PGHVPPGPPGQPPGPAISA----- 71  
Db 123 NCPKGDIGPAGLPGRGPPGPDPGIPGPAGISVGRKGPQGPTEGPPRGPGEKGSVP 182  
Qy 72 -----VGLKDRGATGE 83  
Db 183 GLNGKGEHGHCTPCRPRGERGLPGPGPTGPPGPPGVGRKGENGLPGQGLKGDQGVPE 242  
Qy 84 RGLAGLPQGPSPGPPGPPGPGYGMKATGPMGQGIPIGI---PGPPGPMQGPAGKHCNPS 140  
Db 243 RGAAGSGQGSGGPPGEGPPGIGPAGLPGLPGQGLTGMKGPQGPAPGTAGLPAGPGCKPG 302  
Qy 141 DCFGAMPMEQGYPPMKTMGKPPG 163  
Db 303 -----LPGLKGGQGPVG 314

RESULT 8  
CSB07S  
collagen alpha 1(III) chain - bovine  
C:Species: Bos primigenius taurus (cattle)  
C>Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 09-Jul-2004  
C:Accession: A02862; A38001; A38002; A38003; A38004; A38005; S71946  
R:Fietzek, P.P.; Allmann, H.; Rauterberg, J.; Henkel, W.; Wachter, E.; Kuehn, K.  
Hoppe-Seyler's Z. Physiol. Chem. 360, 809-820, 1979  
A>Title: The covalent structure of calf skin type III collagen. I. The amino acid sequen  
A:Reference number: A02862; MUID:80026026; PMID:488906  
A:Accession: A02862  
A:Molecule type: protein  
A:Residues: 1-242 <FIE>  
A:Cross-references: UNIPROT:P04258; UNIPARC:UPI0000173B8A  
R:Dewes, H.; Fietzek, P.P.; Kuehn, K.

Hoppe-Seyler's Z. Physiol. Chem. 360, 821-832, 1979  
A>Title: The covalent structure of calf skin type III collagen. II. The amino acid sequen  
A:Reference number: A38001; MUID:80026027; PMID:488907  
A:Accession: A38001  
A:Molecule type: protein  
A:Residues: 243-422 <DEWI>  
A:Cross-references: UNIPARC:UPI0000173B8B  
R:Benz, H.; Fietzek, P.P.; Kuehn, K.  
Hoppe-Seyler's Z. Physiol. Chem. 360, 833-840, 1979  
A>Title: The covalent structure of calf skin type III collagen. III. The amino acid sequen  
A:Reference number: A38002; MUID:80026028; PMID:488908  
A:Accession: A38002  
A:Molecule type: protein  
A:Residues: 423-571 <BEN>  
A:Cross-references: UNIPARC:UPI0000173B8C  
R:Yang, H.; Glanville, R.W.; Fietzek, P.P.; Kuehn, K.  
Hoppe-Seyler's Z. Physiol. Chem. 360, 841-850, 1979  
A>Title: The covalent structure of calf skin type III collagen. IV. The amino acid sequen  
A:Reference number: A38003; MUID:80026029; PMID:488909  
A:Accession: A38003  
A:Molecule type: protein  
A:Residues: 572-808 <LAN>  
A:Cross-references: UNIPARC:UPI0000173B8D  
R:Dewes, H.; Fietzek, P.P.; Kuehn, K.  
Hoppe-Seyler's Z. Physiol. Chem. 360, 851-860, 1979  
A>Title: The covalent structure of calf skin type III collagen. V. The amino acid sequen  
A:Reference number: A38004; MUID:80026030; PMID:488910  
A:Accession: A38004  
A:Molecule type: protein  
A:Residues: 809-947 <DEW2>  
A:Cross-references: UNIPARC:UPI0000173B8E  
R:Allmann, H.; Fietzek, P.P.; Glanville, R.W.; Kuehn, K.  
Hoppe-Seyler's Z. Physiol. Chem. 360, 861-868, 1979  
A>Title: The covalent structure of calf skin type III collagen. VI. The amino acid sequen  
A:Reference number: A38005; MUID:80026031; PMID:488911  
A:Accession: A38005  
A:Molecule type: protein  
A:Residues: 948-1049 <ALL>  
A:Cross-references: UNIPARC:UPI0000173B8F  
A:Experimental source: skin  
R:Henkel, W.  
Biochem. J. 318, 497-503, 1996  
A>Title: Cross-link analysis of the C-telopeptide domain from type III collagen.  
A:Reference number: S71946; MUID:96404897; PMID:8809038  
A:Accession: S71946  
A:Molecule type: protein  
A:Residues: 87-106;1017-1029;1037-1049 <HEN>  
A:Cross-references: UNIPARC:UPI0000173B90; UNIPARC:UPI0000173B91; UNIPARC:UPI0000173B92  
C:Comment: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are by  
C:Superfamily: collagen alpha 1(III) chain; fibrillar collagen carboxyl-terminal homology;  
C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyprolin  
F:1-1049/Product: collagen alpha 1(III) chain #status experimental <CAB>  
F:1-14/Region: amino-terminal nonhelical telopeptide  
F:15-1040/Region: helical  
F:587-589/Region: cell attachment (R-G-D) motif  
F:752-754/Region: cell attachment (R-G-D) motif  
F:875-877/Region: cell attachment (R-G-D) motif  
F:878-880/Region: cell attachment (R-G-D) motif  
F:935-937/Region: cell attachment (R-G-D) motif  
F:1041-1049/Region: carboxyl-terminal nonhelical telopeptide  
F:95,107,119,938,950/Modified site: 5-hydroxylysine (lys) #status experimental  
F:107,950/Modified site: allysine (lys) #status predicted  
F:107/Binding site: carbonyl (lys) (covalent) #status experimental  
F:1040,1041/Disulfide bonds: interchain #status predicted

Query Match 40.2%; Score 382.5; DB 1; Length 1049;  
Best Local Similarity 37.2%; Pred. No. 2.5e-17;  
Matches 90; Conservative 9; Mismatches 56; Indels 87; Gaps 6;

Qy 2 GPPGFKGKTGHPGLDPPKGDCKGPPGPPGTPGPAAGEPPGAMP-----GGRPPGPHV 55  
Db 27 GPPGPPGTSGHGPAAGAPGPGPPGPPGQAGPAGPPGPAIGSGKDGSGRPPGPPR 86

QY 56 GPPGPPGPGPAGI-----SAVGLKGD-----R 78  
 Db 87 GPPGPPGKKGKPGAGKPGMGKRGFDGRNGKGGPAGLKGKGVGPDGDGAPGPGMR 146  
 QY 79 GATGRRGLAGLP-----GQPGPPGPGPPGY-----104  
 Db 147 GATGRRGRRGLPGAAGARGNDARSGSDQPGPPGRTAGFGSGPGAKGVEGVPAGSPSS 206  
 QY 105 ---GKMGATGPMGQGGIPIGPPGPPMGQPGKAGHCNPSDCTGAMPEQOYPMKTKMGP 161  
 Db 207 GATGRRGRRGLPGAAGARGNDARSGSDQPGPPGRTAGFGSGPGAKGVEGVPAGSPSS 259  
 QY 162 PG 163  
 Db 260 PG 261  
 RESULT 9  
 CGHUV  
 collagen alpha 1 (V) chain precursor - human  
 N:Alternate names: procollagen alpha 1(V) chain  
 C:Species: Homo sapiens (man)  
 C:Date: 22-Nov-1993 #sequence revision 03-Oct-1995 #text change 09-Jul-2004  
 C:Accession: S18802; S16024; A61142; S11303; S03978; S43642; S58665  
 R:Greenspan, D.S.; Cheng, W.; Hofman, G.G.  
 J: Biol. Chem. 266, 24727-24733, 1991  
 A:Title: The pro-alpha1(V) collagen chain. Complete primary structure, distribution of e  
 A:Reference number: S18802; MUID:92105142; PMID:1722213  
 A:Accession: S18802  
 A:Molecule type: mRNA  
 A:Residues: 1-1838 <G>R>  
 A:Cross-references: UNIPARC:UPI000006E992; GB:M76729; NID:G189519; PIDN:  
 R:Takahara, K.; Sato, Y.; Okawara, K.; Okamoto, N.; Noda, A.; Yaot, Y.; Kato, I.  
 J: Biol. Chem. 266, 13124-13129, 1991  
 A:Title: Complete primary structure of human collagen alpha-1(V) chain.  
 A:Reference number: S16024; MUID:91302336; PMID:2071595  
 A:Accession: S16024  
 A:Molecule type: mRNA  
 A:Residues: 1-81, 'Q', '84-389, 'A', '391-676, 'R', '678-1294, 'PS', '1297, 'RS', '1300-1553, 'R', '1555-  
 A:Cross-references: UNIPARC:UPI0000126DID; GB:D90279; NID:G219509; PIDN:BA414333.1; PID:  
 A:Note: parts of this sequence were determined by protein sequencing  
 R:Yaot, Y.; Hashimoto, K.; Takahara, K.; Kato, I.  
 Exp. Cell Res. 194, 180-185, 1991  
 A:Title: Insulin binds to type V collagen with retention of mitogenic activity.  
 A:Reference number: A61142; MUID:91224163; PMID:1709100  
 A:Accession: A61142  
 A:Molecule type: protein  
 A:Residues: 823-824, 'X', '826-842 <Y>A>  
 A:Cross-references: UNIPARC:UPI0000173BB7  
 A:Note: the residue designated 'X' is probably glycosylated hydroxylysine; this cyanogen  
 R:Yaot, Y.; Hashimoto, K.; Koitabashi, H.; Takahara, K.; Ito, M.; Kato, I.  
 Biochim. Biophys. Acta 1035, 139-145, 1990  
 A:Title: Primary structure of the heparin-binding site of type V collagen.  
 A:Reference number: S11303; MUID:90366601; PMID:2203476  
 A:Accession: S11303  
 A:Molecule type: protein  
 A:Residues: 823-824, 'X', '826-848, 'I', '850-851, 'P', '853, 'PR', '856-893, 'D', '895-932, 'X', '934-950  
 A:Cross-references: UNIPARC:UPI0000173BB8  
 A:Note: the residues designated 'X' are probably glycosylated hydroxylysine; this sequen  
 R:Sever, J.M.; Kang, A.H.  
 Arch. Biochem. Biophys. 271, 120-129, 1989  
 A:Title: Covalent structure of collagen amino acid sequence of three cyanogen bromide-d  
 A:Reference number: S03978; MUID:89227189; PMID:2496661  
 A:Accession: S03978  
 A:Molecule type: protein  
 A:Residues: 621-640, 'G', '642-649, 'L', '651-662, 'E', '664-667, 'Q', '669-676, 'Q', '678-683, 'P', '685-  
 <SE>  
 A:Cross-references: UNIPARC:UPI0000173BB9  
 A:Note: there are a number of inconsistencies between the sequences in figures 6 and 7,  
 R:Morzidi-Ameli, M.; Rousseau, J.C.; Klemm, J.P.; Champilaud, M.F.; Bouillon, M.M.; Ben  
 Eur. J. Biochem. 221, 987-995, 1994  
 A:Title: Diversity in the processing events at the N-terminus of type-V collagen.

A:Reference number: S43642; MUID:94237164; PMID:8181482  
 A:Accession: S43642  
 A:Molecule type: protein  
 A:Residues: 565-576, '756-758, 'X', '760-763, 'X', '765-772, '1012-1029, '1219-1232, '1465-1474, 'X', '14'  
 A:Cross-references: UNIPARC:UPI0000173BB4; UNIPARC:UPI0000173BBB; UNIPARC:UPI0000173BBB;  
 R:Feesler, L.I.; Brosh, S.; Chapin, S.; Feesler, J.H.  
 J: Biol. Chem. 261, 5034-5040, 1986  
 A:Title: Tyrosine sulfation in precursors of collagen V.  
 A:Reference number: A56977; MUID:86168226; PMID:3082875  
 A:Contents: annotation; identification of tyrosine sulfate in the amino-terminal propept  
 R:Lee, S.; Greenspan, D.S.  
 Biochem. J. 310, 15-22, 1995  
 A:Title: Transcriptional promoter of the human alpha-1(V) collagen gene (COL5A1).  
 A:Reference number: S58665; MUID:95374437; PMID:7646438  
 A:Accession: S58665  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 1-36 <LE>  
 A:Cross-references: UNIPARC:UPI00007363D; GB:L38808; NID:G1020325; PIDN:AAA7953.1; PID:  
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C  
 are 5-hydroxylated and subsequently O-glycosylated.  
 C:Comment: A long form of the mature protein containing part of the amino-terminal propep  
 lie the heterotrimers are probably processed to the long form.  
 C:Genetics:  
 A:Gene: GDB:COL5A1  
 A:Cross-references: GDB:131457; OMIM:120215  
 A:Map position: 9q34.2-9q34.3  
 C:Complex: type V collagen may be a homotrimer of alpha 1(V) chains, a heterotrimer of ty  
 alpha 2(V) chain and one alpha 3(V) chain, initially linked by disulfide bonds among the  
 length, is formed with desmosome cross-links made from lysine and allysine residues  
 C:Function:  
 A:Description: structural component of extracellular fibrous polymer associated with cell  
 A:Note: may play a role in controlling the lateral growth of collagen I fibrils  
 C:Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology  
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyprolin  
 F:1-37/Domain: signal sequence #status predicted <S>G>  
 F:136-261/Domain: PARP-like #status predicted <PAP>  
 F:138-541/Domain: amino-terminal propeptide #status predicted <PRO>  
 F:542-1605/Product: collagen alpha 1(V) chain, short form #status predicted <MAT>  
 F:542-558/Region: amino-terminal nonhelical telopeptide  
 F:559-1572/Region: helical  
 F:1645-647/Region: cell attachment (R-G-D) motif  
 F:663-665/Region: cell attachment (R-G-D) motif  
 F:897-929/Region: heparin binding  
 F:1573-1605/Region: carboxyl-terminal nonhelical telopeptide  
 F:1606-1838/Domain: carboxyl-terminal propeptide #status predicted <CPR>  
 F:1615-1837/Domain: fibrillar collagen carboxyl-terminal homology <FCC>  
 F:38/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted  
 F:62-244, 183-237/Disulfide bonds: #status predicted  
 F:159, 176, 385, 1672, 1741/Binding site: carboxylate (Asn) (covalent) #status predicted  
 F:234, 236, 240, 262, 263, 273, 274, 275, 277, 279, 280, 338, 340, 346, 347, 352, 357, 416, 417, 420, 421/Bit  
 F:535/Modified site: allysine (Lys) #status predicted  
 F:541-542/Cleavage site: Ala-Gln (procollagen N-endopeptidase) #status predicted  
 F:542/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted  
 F:570, 576, 621, 639, 648, 654, 657, 675, 678, 690, 693, 696, 705, 717, 720, 726, 732, 741, 750, 753, 756, 76;  
 site: 4-hydroxyproline (Pro) #status experimental  
 F:627, 642, 687, 708, 744, 774, 795, 804, 807, 810, 819, 825, 846, 864, 882, 897/Modified site: 5-hydro  
 F:627, 642, 687, 774, 795, 804, 807, 810, 819, 825, 846, 864, 882, 897, 1482/Binding site: carboxylate  
 F:708, 744/Binding site: carboxylate (Lys) (covalent) #status experimental  
 F:1482/Modified site: 5-hydroxylysine (Lys) #status predicted  
 F:1605-1606/Cleavage site: Ala-Asp (procollagen C-endopeptidase) #status predicted  
 F:1639, 1645, 1662, 1671/Disulfide bonds: interchain #status predicted  
 F:1680-1835, 1746-1789/Disulfide bonds: #status predicted  
 Query Match 40.2%; Score 382.5; DB 1; Length 1838;  
 Best Local Similarity 39.7%; Pred. No. 4e-11;  
 Matches 83; Conservative 19; Mismatches 56; Indels 51; Gaps 4;  
 QY 2 GPPGPPGKKGKPGAGKPGMGKRGFDGRNGKGGPAGLKGKGVGPDGDGAPGPGMR 61  
 Db 1291 GPPGPPGKKGKPGAGKPGMGKRGFDGRNGKGGPAGLKGKGVGPDGDGAPGPGMR 1350  
 QY 62 GPPGPPGAG-----ISAVGLKGDGRAT 81





A>Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide sequence.  
 R.Seyer, J.M.; Kang, A.H.  
 Biochemistry 16, 1158-1164, 1977  
 A>Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptide  
 A:Reference number: A90399; PMID:77134724; PMID:557335  
 A:Accession: A90399  
 A:Molecule type: protein  
 A:Residues: 'V', 169-225, 229-232, 'P', 234-292, 'D', 294-398 <SEV1>  
 A:Cross-references: UNIPARC:UPI0000173881  
 A:Experimental source: liver  
 A>Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactosyl  
 R.Seyer, J.M.  
 submitted to the Atlas, December 1977  
 A:Reference number: A94562  
 A:Accession: A94562  
 A:Molecule type: protein  
 A:Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEV2>  
 A:Cross-references: UNIPARC:UPI0000173882  
 A:Experimental source: liver  
 A>Note: author submitted corrections to A90399  
 R.Milwicz, D.M.; Wiltz, A.M.; Smith, A.C.; Manchester, D.K.; Walderstein, G.; Byers, P.H.  
 Am. J. Hum. Genet. 53, 62-70, 1993  
 A>Title: Parental somatic and germ-line mosaicism for a multigene deletion with unusual  
 karyotype.  
 A:Reference number: 151868; PMID:93304430; PMID:8317500  
 A:Accession: 151868  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 186-194 <MTL>  
 A:Cross-references: UNIPARC:UPI0000000B14; GB:S62925; NID:G386425; PIDN:AA13937.1; PID:  
 R.Chiodo, A.A.; Silience, D.O.; Cole, W.G.; Bateman, J.F.  
 Biochem. J. 311, 939-943, 1995  
 A>Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL3A1  
 A:Reference number: 559511; PMID:96057614; PMID:7487954  
 A:Accession: 559511  
 A:Molecule type: mRNA  
 A:Residues: 302-423 <CHI>  
 A:Cross-references: UNIPARC:UPI0000173883; GB:S79877; NID:G1195576; PIDN:AA35615.1; PID:  
 R.Seyer, J.M.; Kang, A.H.  
 Biochemistry 17, 3404-3411, 1978  
 A>Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr pe  
 A:Reference number: A90414; PMID:79000343; PMID:687591  
 A:Accession: A90414  
 A:Molecule type: protein  
 A:Residues: 399-675, 'N', 677-727 <SEV3>  
 A:Cross-references: UNIPARC:UPI0000173884  
 A:Experimental source: liver  
 R.Lee, B.; Vitale, E.; Superfi-Furga, A.; Steimann, B.; Ramirez, F.  
 J. Biol. Chem. 266, 5256-5259, 1991  
 A>Title: G to T transversion at position +5 of a splice donor site causes skipping of th  
 A:Reference number: 155349; PMID:9116161; PMID:1672129  
 A:Accession: 155349  
 A:Molecule type: protein  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 537-605 <LEE>  
 A:Cross-references: UNIPARC:UPI0000000442; GB:M59312; NID:G180815; PIDN:AA52041.1; PID:  
 R.Seyer, J.M.; Mainardi, C.; Kang, A.H.  
 Biochemistry 19, 1583-1589, 1980  
 A>Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from ty  
 A:Reference number: A90438; PMID:80198282; PMID:6246925  
 A:Accession: A90438  
 A:Molecule type: protein  
 A:Residues: 728-895, 'A', 897-964 <SEV4>  
 A:Cross-references: UNIPARC:UPI0000173885  
 A:Experimental source: liver  
 R.Cole, W.G.; Chiodo, A.A.; Lemande, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Chan  
 J. Biol. Chem. 265, 17070-17077, 1990  
 A>Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping an  
 A:Reference number: A38303; PMID:91009133; PMID:2145268  
 A:Accession: A38303  
 A:Molecule type: mRNA  
 A:Residues: 861-1015 <COL>  
 A:Cross-references: UNIPARC:UPI0000000441; GB:S05617; GB:M55603; GB:M59227; NID:G180878;

A>Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos syn  
 R.Mankoo, B.S.; Dalgleish, R.  
 Nucleic Acids Res. 16, 2337, 1988  
 A>Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.  
 A:Reference number: 802119; PMID:88189827; PMID:3357782  
 A:Accession: 802119  
 A:Molecule type: mRNA  
 A:Status: translation not shown  
 A:Residues: 950-1018, 'Y', 1020-1183, 'S', 1185-1466 <MAN>  
 A:Cross-references: UNIPARC:UPI0000173886; EMBL:X06700; NID:G30053; PIDN:CAA29886.1; PID:  
 R.Seyer, J.M.; Kang, A.H.  
 Biochemistry 20, 2621-2627, 1981  
 A>Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from ty  
 A:Reference number: A90446; PMID:81208139; PMID:7016180  
 A:Accession: A90446  
 A:Molecule type: protein  
 A:Residues: 965-979, 'A', 981-984, 'PS', 987, 'QN', 990-1096, 'P', 1098-1152, 'AT', 1155, 'S', 1157-1  
 A:Cross-references: UNIPARC:UPI0000173887  
 A:Experimental source: liver  
 R.Loidi, H.R.; Brinker, J.M.; May, M.; Phlajant, T.; Morrow, S.; Rosenbloom, J.; Mye  
 Nucleic Acids Res. 12, 9383-9394, 1984  
 A>Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollagen  
 A:Reference number: A93551; PMID:85087944; PMID:6096827  
 A:Accession: A93551  
 A:Molecule type: mRNA  
 A:Residues: 1065-1155, 'P', 1157-1466 <LOI>  
 A:Cross-references: UNIPARC:UPI000016A6B5; GB:M13146; NID:G180415; PIDN:AA52003.1; PID:  
 R.Mankoo, B.S.; Dalgleish, R.; Kluge-Beckman, B.; Renard, S.I.; Tolstoev, F.; Brant  
 Biochemistry 25, 1408-1413, 1986  
 A>Title: Human type III collagen gene expression is coordinately modulated with the type  
 A:Reference number: 152393; PMID:86187804; PMID:3754462  
 A:Accession: 152393  
 A:Molecule type: protein  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1161-1200 <MIS>  
 A:Cross-references: UNIPARC:UPI000016A6B5; GB:M13146; NID:G180415; PIDN:AA52003.1; PID:  
 R.Emanuel, B.S.; Camizato, L.A.; Seyer, J.M.; Myers, J.C.  
 Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985  
 A>Title: Isolation of cDNA and genomic clones encoding human pro-alpha1(III) collagen. Pe  
 A:Reference number: 159025; PMID:85157600; PMID:2579949  
 A:Accession: 159025  
 A:Molecule type: DNA  
 A:Residues: 1176-1240, 'V', 1242-1356, 'P', 1358-1466 <CHU>  
 A:Cross-references: UNIPARC:UPI0000173889; GB:M10615; GB:M10793; GB:M10794; GB:M10795; G  
 A:Experimental source: liver  
 A>Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given for  
 action  
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C  
 3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently O-  
 C:Genetics:  
 A:Gene: GDB:COL3A1  
 A:Cross-references: GDB:118729; OMIM:120180  
 A:Map position: 2q31-2q31  
 A:Introns: 27/1, 94/3, 111/3, 149/3, 176/3, 554/3, 587/3, 1175/3, 1275/1, 1337/3, 1418/3  
 A>Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Dan  
 C:Complex: type III collagen is a homotrimer of monomers initially linked by disulfide b  
 er of their length, is formed with desmosine cross-links made from lysine and allylsine re  
 C:Function:  
 A:Description: structural component of extracellular fibrous polymer that maintains intec  
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
 C:Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hyd  
 F/1-23/Domain: signal sequence #status predicted <SID>  
 F/24-153/Domain: amino-terminal propeptide #status predicted <PRO>  
 F/31-91/Domain: von Willebrand factor type C repeat homology <VWC>

F,154-1221/Product: collagen alpha 1(III) chain #status predicted <Mat>  
 F,154-167/Region: amino-terminal nonhelical telopeptide  
 F,168-1196/Region: helical  
 F,1091-1093/Region: cell attachment (R-G-D) motif  
 F,1197-1221/Region: carboxyl-terminal nonhelical telopeptide  
 F,1222-1466/Domain: carboxyl-terminal propeptide #status predicted <CPR>  
 F,1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <FCC>  
 F,124/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted  
 F,153-154/Cleavage site: Pro-Gln (procollagen N-endopeptidase) #status predicted  
 F,154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted  
 F,161,1212/Modified site: allysine (lys) #status predicted

Query Match 39.6%; Score 377; DB 1; Length 1466;  
 Best local Similarity 35.0%; Pred. No. 7,4e-17;  
 Matches 90; Conservative 8; Mismatches 57; Indels 102; Gaps 6;

QY	2	GPPGFKGKTGHPGLPGPKGDCGKPPPGSTGRPGAEGEPGAMGPOGRPPGHVGP----	57
Db	168	GLAGYPRAPGRPPGPPPTSGHPGSPGSGPYQGPGEQAGPSGPPGALGPSGPA	227
QY	58	-----PGRPGQPPGAGI-----	77
Db	228	GKDGEGRPRPGERGLPPGPGIKGPAGI.PGPPGKMGHRGPDGRNGEKGTGAPGKGEN	287
QY	78	-----RGATGERGLAGLP-----	104
Db	288	GLPGENGAPGPMGRKAPGSRGRRGLPGAAGARNDGARGSDQPPGPPPTAGFPGSP	347
QY	105	-----GKMGATGPMGQOQI.PGI.PGPPGPMGQPPGKAGHCNPSDCFGAM	146
Db	348	GAKGEVPAGSPGSGNAPGQGRBEPGQHAAGGPPGPPGINSFGKGMGMPAGIPGA-	406
QY	147	PMEQOYPPMKTKMGPPG	163
Db	407	-----PGLMGARGPPG	417

Search completed: December 1, 2005, 08:27:06  
 Job time : 18 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2005, 08:15:46 ; Search time 228 Seconds  
(without alignments)  
504.391 Million cell updates/sec

Title: US-09-924-340-58  
Perfect score: 952  
Sequence: 1 MGPFGFKGTGHGRLPGPKG.....GAMPMEQGYRPMKTMKAPFG 163

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : UniProt\_05.80:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	840.5	88.3	1463	2	Q59F89_HUMAN
2	840.5	88.3	1603	1	COGAL_HUMAN
3	800.5	84.1	575	2	Q8CIF9_MOUSE
4	794.5	83.5	415	2	Q568Y4_RAT
5	777	81.6	1560	2	Q8BLX7_MOUSE
6	620	65.1	218	2	Q71RG9_HUMAN
7	558	58.6	150	2	Q9CZS2_MOUSE
8	407	42.8	925	2	Q4SIU4_TETNG
9	391	41.1	799	2	Q8BNS7_MOUSE
10	381	41.1	1691	2	Q9ESQ2_MOUSE
11	391	41.1	1691	2	Q632W6_MOUSE
12	390	41.0	1685	1	CO4A5_HUMAN
13	390	41.0	1685	1	Q9NIB7_HUMAN
14	387	40.7	1146	1	COHAL_CHICK
15	387	40.7	1453	1	COHAL_CHICK
16	385	40.4	1497	1	COHAL_HUMAN
17	384	40.3	1549	2	Q60444_CRIGR
18	384	40.3	1748	2	Q591P1_PIG
19	383.5	40.3	1840	2	CO5A1_HUMAN
20	383.5	40.3	1840	2	Q591P3_PIG
21	383	40.2	355	2	Q5YFPG_YVIRU
22	383	40.2	674	1	COA1_BOVIN
23	382.5	40.2	591	2	Q96HC0_HUMAN
24	382.5	40.2	848	2	Q4S270_TETNG
25	382.5	40.2	1049	1	CO3A1_BOVIN
26	382.5	40.2	1415	2	Q5JVJ6_HUMAN
27	382.5	40.2	1792	2	Q59SE7_HUMAN
28	382.5	40.2	1838	2	Q15094_HUMAN
29	382.5	40.2	1838	2	Q5SUX4_HUMAN
30	382.5	40.2	1838	2	Q88207_MOUSE
31	382.5	40.2	1840	2	Q60467_CRILLO

32	382.5	40.2	1840	2	Q9J103_RAT	Q9J103_rattus norv
33	382	40.1	1694	2	Q8HYC1_CANFA	Q8HYC1_canis fam1
34	382	40.1	1691	1	CO4A5_CANFA	Q28247_canis fam1
35	381.5	40.1	657	2	Q4VAP9_HUMAN	Q4VAP9_homo sapien
36	381.5	40.1	703	1	CO8A2_HUMAN	P25067_homo sapien
37	381.5	40.1	703	2	Q5VJ31_HUMAN	Q5VJ31_homo sapien
38	381	40.0	367	2	Q5GAF3_YVIRU	Q5GAF3_groupfer iri
39	380.5	40.0	1470	1	CO5A3_HUMAN	Q07563_mus musculu
40	380.5	40.0	1835	2	Q9IAU4_CHICK	Q9IAU4_gallus gall
41	380	39.9	1806	2	Q5VJ31_HUMAN	Q5VJ31_homo sapien
42	379	39.8	675	2	Q9N178_PIG	Q9N178_sus scrofa
43	378.5	39.8	1269	2	Q7T227_CHICK	Q7T227_gallus gall
44	378.5	39.8	1658	2	Q59GD4_HUMAN	Q59GD4_homo sapien
45	378.5	39.8	1745	1	CO5A3_HUMAN	P25940_homo sapien
46	378	39.7	1445	2	Q93251_RANCA	Q93251_rana catesb
47	377.5	39.7	888	2	Q90796_CHICK	Q90796_gallus gall
48	377	39.6	568	2	Q8CD80_MOUSE	Q8CD80_mus musculu
49	377	39.6	1163	2	Q8N6U4_HUMAN	Q8N6U4_homo sapien
50	377	39.6	1463	2	Q5PQT6_RAT	Q5PQT6_rattus norv
51	377	39.6	1463	1	CO3A1_HUMAN	P02461_homo sapien
52	377	39.6	1466	2	Q53S91_HUMAN	Q53S91_homo sapien
53	376.5	39.5	1466	2	Q541P8_HUMAN	Q541P8_homo sapien
54	376.5	39.5	1449	2	Q6NZ15_BRARE	Q6NZ15_brachydanio
55	376.5	39.5	1487	2	Q77J53_CANFA	Q77J53_canis fam1
56	376.5	39.5	1690	1	CO4A4_HUMAN	P53420_homo sapien
57	376.5	39.5	1690	1	Q53WRI_HUMAN	Q53WRI_homo sapien
58	376	39.5	1652	1	CO3A1_CHICK	P12105_gallus gall
59	375	39.5	1450	2	Q9Y1B4_CYNPY	Q9Y1B4_cynops pyrr
60	375.5	39.4	1449	2	Q6PE19_BRARE	Q6PE19_brachydanio
61	374.5	39.3	585	2	Q80V57_MOUSE	Q80V57_mus musculu
62	374	39.3	680	2	Q9D0D2_MOUSE	Q9D0D2_m_mus muscu
63	374	39.3	1464	1	CO3A1_MOUSE	P08121_mus musculu
64	374	39.3	1464	2	Q8BKX2_MOUSE	Q8BKX2_mus musculu
65	374	39.3	1464	2	Q8BLW4_MOUSE	Q8BLW4_mus musculu
66	374	39.3	1464	2	Q7TT32_MOUSE	Q7TT32_mus musculu
67	373.5	39.2	1467	2	Q5DTG2_MOUSE	Q5DTG2_mus musculu
68	373.5	39.2	1804	1	COBA1_MOUSE	Q61245_mus musculu
69	373.5	39.2	1804	1	Q80WR4_MOUSE	Q80WR4_mus musculu
70	373	39.2	890	2	Q77087_YANNE	Q77087_alvinella p
71	373	39.2	1770	2	Q701V4_ANOGA	Q701V4_anopheles g
72	372.5	39.1	280	2	Q91014_CHICK	Q91014_gallus gall
73	372.5	39.1	680	1	COA1_HUMAN	Q03622_homo sapien
74	372.5	39.1	739	2	Q70575_MOUSE	Q70575_mus musculu
75	372.5	39.1	977	2	Q4SEP8_TETNG	Q4SEP8_tetraodon n
76	372.5	39.1	1420	2	Q90W37_CHICK	Q90W37_gallus gall
77	372	39.1	662	2	Q8BXY3_MOUSE	Q8BXY3_mus musculu
78	372	39.1	680	2	Q8BSX1_MOUSE	Q8BSX1_mus musculu
79	372	39.1	920	1	CO9A1_CHICK	P12106_gallus gall
80	372	39.1	1069	2	Q6LAN8_HUMAN	Q6LAN8_homo sapien
81	372	39.1	1449	2	Q640B2_XENTR	Q640B2_xenopus tro
82	372	39.1	1461	2	Q76045_HUMAN	Q76045_homo sapien
83	372	39.1	1464	1	COA1_HUMAN	P02452_homo sapien
84	372	39.1	1464	2	Q8N473_HUMAN	Q8N473_homo sapien
85	372	39.1	1467	2	Q59F64_HUMAN	Q59F64_homo sapien
86	371.5	39.0	730	2	Q26052_PARLI	Q26052_parencetrot
87	371.5	39.0	1017	2	Q59HB5_HUMAN	Q59HB5_homo sapien
88	371.5	39.0	1806	1	COBA1_HUMAN	P12107_homo sapien
89	371	39.0	636	1	COA1_RAT	P13911_rattus norv
90	371	39.0	779	1	COA1_BOVIN	P02453_bos taurus
91	371	39.0	1460	1	COA1_CANFA	Q9X677_canis fam1
92	371	39.0	1631	2	Q59700_CANFA	Q59700_canis fam1
93	371	39.0	1707	1	CO4A2_MOUSE	P08122_mus musculu
94	370.5	38.9	826	2	Q8KON6_MOUSE	Q8KON6_mus musculu
95	370.5	38.9	886	2	Q8CEP7_MOUSE	Q8CEP7_mus musculu
96	370.5	38.9	886	2	Q80VY3_MOUSE	Q80VY3_mus musculu
97	370.5	38.9	1419	2	Q80X38_MOUSE	Q80X38_mus musculu
98	370.5	38.9	1419	2	Q63123_RAT	Q63123_rattus norv
99	370.5	38.9	1442	2	Q62031_MOUSE	Q62031_mus musculu
100	370.5	38.9	1442	2	Q62033_MOUSE	Q62033_mus musculu

## ALIGNMENTS

RESULT 1  
 OS9F89 HUMAN PRELIMINARY; PRT; 1463 AA.  
 ID OS9F89\_ HUMAN  
 AC OS9F89; 10-MAY-2005 (TRENBLREL. 30, Created)  
 DT 10-MAY-2005 (TRENBLREL. 30, Last sequence update)  
 DT 10-MAY-2005 (TRENBLREL. 30, Last sequence update)  
 DE Alpha 1 type XVI collagen variant (Fragment).  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;  
 OC Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Splicein;  
 RA Totoki Y., Toyoda A., Takeda T., Sakaki Y., Tanaka A., Yokoyama S.,  
 RA Ohara O., Nagase T., Kikuno F.R.;  
 RT "None Title";  
 RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB209571; BMD92808.1; -; mRNA.  
 DR GO; GO:0005737; C:cytoplasm; IEA.  
 DR GO; GO:0005578; C:extracellular matrix (sensu Metazoa); IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0007155; P:cell adhesion; IEA.  
 DR GO; GO:006817; P:phosphate transport; IEA.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR003129; Laminin\_G\_TSP\_N.  
 DR Pfam; PF01391; Collagen; 14.  
 DR SMART; SM00210; TSPN; 1.  
 DR KW Collagen; Extracellular matrix.  
 FT NON TER 1  
 SQ SEQUENCE 1463 AA; 142798 MM; 486F0B10CA93785C CRC64;

Query Match 88.3%; Score 840.5; DB 2; Length 1463;  
 Best Local Similarity 44.5%; Pred. No. 3.6e-42;  
 Matches 163; Conservative 0; Mismatches 0; Indels 203; Gaps 1;

1 MPPPGKGTGHPGLPGPKGDCGKPPGSTGRPGAEGPGAMGPGGPGPHVPPGP 60  
 1098 MPPPGKGTGHPGLPGPKGDCGKPPGSTGRPGAEGPGAMGPGGPGPHVPPGP 1157  
 61 PGPAGISAVGLKGDGRATGERGLAGPGQ----- 92  
 1158 PGPAGISAVGLKGDGRATGERGLAGPGQGGPPGHPGPPGPTDGAAGKSPRQ 1217  
 93 ----- 92  
 1218 GFYPPGPKDPPGAAGQKGAEGKRAHMPGPGKSGMGPVGPQPPAGEGHGAPGS 1277  
 93 ----- 92  
 1278 GSPGLPGVPSMGDMVYDEIKRIFRQEI11KMFDEMAVYTSRQPMEMAAAQGRPGPP 1337  
 93 ----- -GPPG 97  
 1338 GKDGAAGRGARQSPGLPQIGRGROGLPVGRLPDTGKEDDITG1AGENGGLPDPG 1397  
 98 PGPAGISAVGLKGDGRATGERGLAGPGQGGPPGHPGPPGPTDGAAGKSPRQ 157  
 1398 PGPAGISAVGLKGDGRATGERGLAGPGQGGPPGHPGPPGPTDGAAGKSPRQ 1457  
 158 MKGPPG 163  
 1458 MKGPPG 1463

RESULT 2  
 COGAL HUMAN STANDARD; PRT; 1603 AA.  
 AC COG07092;  
 DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 10-MAY-2005 (Rel. 47, Last annotation update)  
 DE Collagen alpha 1(XVI) chain precursor.  
 GN Name=COL16A1;  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;  
 OC Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=9235339; PubMed=161157;  
 RA Pan T.C., Zhang R.Z., Mattei M.-G., Timpl R., Chu M.-L.,  
 RT "Cloning and chromosomal location of human alpha 1(XVI) collagen.",  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:6565-6569(1992).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE OF 418-1603.  
 RC TISSUE=Placenta;  
 RX MEDLINE=93203161; PubMed=1284248;  
 RA Yamaguchi N., Kimura S., McBride O.W., Hori H., Yamada Y.,  
 RA Kanamori T., Yamakoshi H., Nagai Y.;  
 RT "Molecular cloning and partial characterization of a novel collagen  
 chain, alpha 1(XVI), consisting of repetitive collagenous domains and  
 cysteine-containing non-collagenous segments."  
 RL J. Biochem. 112:856-863(1992).  
 CC -1- FUNCTION: The numerous interruptions in the triple helix may make  
 this molecule either elastic or flexible.  
 CC -1- TISSUE SPECIFICITY: In the placenta, where it is found in the  
 amnion, a membranous tissue lining the amniotic cavity. Within the  
 amnion, it is found in an acellular, relatively dense layer of a  
 complex network of reticular fibers. Also located to a fibroblast  
 layer beneath this dense layer. Exists in tissues in association  
 with other types of collagen.  
 CC -1- DEVELOPMENTAL STAGE: Transiently elevated expression during  
 gestation, and decrease at term.  
 CC -1- DOMAIN: This sequence defines eighteen different domains, nine  
 triple-helical domains (COL9 to COL1) and ten nontriple-helical  
 domains (NC10 to NC1).  
 CC -1- PTM: Prolines at the third position of the tripeptide repeating  
 unit (G-X-Y) are hydroxylated in some or all of the chains.  
 CC -1- SIMILARITY: Belongs to the fibril-associated collagens with  
 interrupted helices (FACIT) family.  
 CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use as long as its content is in no way modified and this statement is not  
 removed.  
 CC EMBL; M92642; AAA58427.1; -; mRNA.  
 CC EMBL; S57132; AAB25797.1; -; mRNA.  
 DR PIR; S23810; S23810.  
 DR Ensembl; ENSG0000084636; Homo sapiens.  
 DR HGNC; HGNC:2193; COL16A1.  
 DR MIM; 120326; -.  
 DR GO; GO:0005597; P:collagen type XVI; TAS.  
 DR GO; GO:0007565; P:pregnancy; TAS.  
 DR InterPro; IPR008161; C1g\_helix.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR003129; Laminin\_G\_TSP\_N.  
 DR Pfam; PF01391; Collagen; 16.  
 DR ProDom; PD0000007; C1g\_helix; 1.  
 DR SMART; SM00210; TSPN; 1.  
 KW Collagen; Extracellular matrix; Hydroxylation; Repeat; Signal;  
 KW Structural protein.  
 FT SIGNAL 1 21 Potential.  
 FT CHAIN 22 1603 Collagen alpha 1(XVI) chain.  
 FT DOMAIN 50 231 TSP N-terminal.  
 FT REGION 232 374 Nonhelical region 10 (NC10).  
 FT REGION 375 505 Triple-helical region 9 (COL9) with 3  
 FT FT imperfections.  
 FT REGION 506 520 Nonhelical region 9 (NC9).

FT REGION 521 554 Triple-helical region 8 (COL8) with 1  
 FT REGION 555 571 Imperfection.  
 FT REGION 572 630 Nonhelical region 8 (NC8).  
 FT REGION 631 651 Triple-helical region 7 (COL7) with 1  
 FT REGION 652 722 Imperfection.  
 FT REGION 723 737 Nonhelical region 7 (NC7).  
 FT REGION 736 875 Triple-helical region 6 (COL6) with 1  
 FT REGION 876 886 Imperfection.  
 FT REGION 887 938 Nonhelical region 5 (NC5).  
 FT REGION 939 972 Triple-helical region 4 (COL4) with 2  
 FT REGION 973 987 Imperfection.  
 FT REGION 988 1010 Nonhelical region 3 (NC3).  
 FT REGION 1011 1432 Triple-helical region 2 (COL2) with 2  
 FT REGION 1433 1471 Imperfection.  
 FT REGION 1472 1577 Nonhelical region 1 (NC1).  
 FT REGION 1578 1603 Triple-helical region 1 (COL1) with 2  
 FT CONFLICT 418 420 RDA -> GGR (in Ref. 2).  
 FT CONFLICT 537 537 R -> P (in Ref. 2).  
 FT CONFLICT 1160 1160 T -> P (in Ref. 2).  
 FT CONFLICT 1163 1163 T -> P (in Ref. 2).  
 FT CONFLICT 1165 1165 S -> P (in Ref. 2).  
 SQ SEQUENCE 1603 AA; 157693 MW; E27D9A1D4E598A37 CRC64;

Query Match 88.3%; Score 840.5; DB 1; Length 1603;  
 Best Local Similarity 44.5%; Pred. No. 3.9e-42;  
 Matches 163; Conservative 0; Mismatches 0; Indels 203; Gaps 1;

QY 1 MGPPGKGTGTHGRLPGPKDCKPSPGSGTGRPAEGEPGAMGPGRPGPHVGP 60  
 DB 1238 MGPPGKGTGTHGRLPGPKDCKPSPGSGTGRPAEGEPGAMGPGRPGPHVGP 1297  
 QY 61 PGPGPAGISAVGLKDRGATGSRGLAGLPGO----- 92  
 DB 1298 PGPGPAGISAVGLKDRGATGSRGLAGLPGO----- 92  
 QY 93 ----- 92  
 DB 1358 GFYGPSPKGDPAAGKQAGKGRAGMRGKSGSMGVPVPPGABGRHGPAGPS 1417  
 QY 93 ----- 92  
 DB 1418 GSPGLPGVDSMGDMVNDIKRIFRQELIKMFDERMAYTYSHPFMEVAAAPGRBPP 1477  
 QY 93 -----PGPRG 97  
 DB 1478 GKDGAAPRPAAGPSGLPGQIGREGROGLPGVRLPGTYGKDKDITGAGENGGLPGPPG 1537  
 QY 98 PGPPGPGATGATGPMGQGGIGPPIPGPPGPMGQPGKAGHCNPSDCGAMPMEQOYPPMKT 157  
 DB 1538 PGPPGPGATGATGPMGQGGIGPPIPGPPGPMGQPGKAGHCNPSDCGAMPMEQOYPPMKT 1597  
 QY 158 MKGPPG 163  
 DB 1598 MKGPPG 1603

RESULT 3  
 Q8CIF9\_MOUSE PRELIMINARY; PRT; 575 AA.  
 AC 08CIF9;  
 DT 01-MAR-2003 (TEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)  
 DE Coll6a1 protein (Fragment).  
 GN Name=Coll6a1;  
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Murine; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP NCLOTIDE SEQUENCE.  
 RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Peingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Stennem C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusik A., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Hellon E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP NCLOTIDE SEQUENCE.  
 RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;  
 RA Strausberg R.;  
 RL Submitted (FEb-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC023940; AAH23940.1; -; mRNA.  
 DR Ensembl: ENSMUSG000000040690; Mus musculus.  
 DR MGI: MGI:1095396; Coll6a1.  
 DR GO: GO:0005737; C:cytoplasm; IEA.  
 DR GO: GO:0006817; P:phosphate transport; IEA.  
 DR InterPro: IPR008161; C1g\_helix.  
 DR InterPro: IPR008160; Collagen.  
 DR Pfam: PF01391; Collagen; 9.  
 DR Prodom: PD000007; C1g\_helix; 1.  
 FT Collagen.  
 FT NON TER 1 1  
 SQ SEQUENCE 575 AA; 54847 MW; 0B5420D04086A13 CRC64;

Query Match 84.1%; Score 800.5; DB 2; Length 575;  
 Best Local Similarity 42.1%; Pred. No. 4e-40;  
 Matches 154; Conservative 4; Mismatches 5; Indels 203; Gaps 1;

QY 1 MGPPGKGTGTHGRLPGPKDCKPSPGSGTGRPAEGEPGAMGPGRPGPHVGP 60  
 DB 210 MGPPGKGTGTHGRLPGPKDCKPSPGSGTGRPAEGEPGAMGPGRPGPHVGP 269  
 QY 61 PGPPGAGISAVGLKDRGATGSRGLAGLPGO----- 92  
 DB 270 PGPPGAGISAVGLKDRGATGSRGLAGLPGO----- 92  
 QY 93 ----- 92  
 DB 330 GLYGPSPKGDPAAGKQAGKGRAGMRGKSGSMGPIGPAPABGRHGPAGPS 389  
 QY 93 ----- 92  
 DB 390 GNPGLPGVDSMGDMVNDIKRIFRQELIKMFDERMAYTYSHPFMEVAAAPGRBPP 449  
 QY 93 -----PGPRG 97  
 DB 450 GKDGAAPRPAAGPSGLPGQIGREGROGLPGVRLPGTYGKDKDITGAGENGGLPGPPG 509  
 QY 98 PGPPGPGATGATGPMGQGGIGPPIPGPPGPMGQPGKAGHCNPSDCGAMPMEQOYPPMKT 157  
 DB 510 PGPPGPGATGATGPMGQGGIGPPIPGPPGPMGQPGKAGHCNPSDCGAMPMEQOYPPMKS 569

```
OY 158 MKGPRG 163
Db 570 MKGPRG 575

RESULT 4
O568Y4 RAT PRELIMINARY; PRT; 415 AA.
ID O568Y4 RAT
AC O568Y4 RAT
DT 10-MAY-2005 (TRENBLREL. 30, Last sequence update)
DT 10-MAY-2005 (TRENBLREL. 30, Last annotation update)
DE Col16a1 predicted protein.
DE Name=Col16a1 predicted.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Rattus.
OC NCBI_Taxid=10116;
[1]
RC NUCLEOTIDE SEQUENCE.
RA STRAUSBERG R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheaffer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant J.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mallaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muray D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton B., Kettman W., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Tuchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smalls D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[2]
RC NUCLEOTIDE SEQUENCE.
RA TISSUE=Brain;
RA NIH MGC Project;
RA Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
RA EMBL; BC092654; AAH92654.1; -; mRNA.
RA GO; GO:0005737; Cytoplasm; IEA.
RA GO; GO:0006817; Phosphate transport; IEA.
RA Collagen.
SQ SEQUENCE 415 AA; 39879 MW; 693F7D948237D1AA CRC64;

Query Match 83.5%; Score 794.5; DB 2; Length 415;
Best Local Similarity 41.5%; Pred. No. 7e-40;
Matches 152; Conservative 7; Mismatches 4; Indels 203; Gaps 1;
```

```
Db 230 GNPGLPGLPGSMGVNVDIDIKRFIRQELIKLFDERMAYYSRMQPMVEVAAAPGRPGP 289
OY 93
Db 290 GKDGA-PGRPGAPSPGLPQIGREGRGGLPGMRGLPGTKGDKDGIQVGIAGENG.LPGPG 349
OY 98 PGPPPGYGMGATGPMGQGGIPGIPGPPGPMQPGKAGHCNPSDCFGAMPMEQOYPPMT 157
Db 350 PGPPPGYGMGATGPMGQGGIPGIPGPPGPMQPGKAGHCNPSDCFGAMPMEQOYPPMX 409
OY 158 MKGPRG 163
Db 410 MKGPRG 415

RESULT 5
O8BLX7 MOUSE
ID O8BLX7 MOUSE PRELIMINARY; PRT; 1580 AA.
AC O8BLX7;
DT 01-MAR-2003 (TRENBLREL. 23, Created)
DT 01-MAR-2003 (TRENBLREL. 23, Last sequence update)
DT 01-MAR-2004 (TRENBLREL. 26, Last annotation update)
DE Mus musculus adult male aorta and vein cDNA. RIKEN full-length
DE enriched library, clone:A53052M23 product:COLLAGEN ALPHA 1(XVI) CHAIN
DE homolog.
DE Name=Col16a1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OC NCBI_Taxid=10090;
[1]
RC NUCLEOTIDE SEQUENCE.
RA STRAIN=C57BL/6J; TISSUE=Aorta and vein;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.,
RT "High-efficiency full-length cDNA cloning."
RT Meth. Enzymol. 303:19-44 (1999).
[2]
RC NUCLEOTIDE SEQUENCE.
RA STRAIN=C57BL/6J; TISSUE=Aorta and vein;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai T., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Glass C., King B., Kochiwa H.,
RA Kueni P., Lewis S., Matsuo T., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.W., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzaletti U., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyokita K., Wang K.H., Weitz C., Whitteker C., Wilming L.,
RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection."
RT Nature 409:685-690 (2001).
[3]
RC NUCLEOTIDE SEQUENCE.
RA STRAIN=C57BL/6J; TISSUE=Aorta and vein;
RA The FANTOM Consortium.
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RT Nature 420:563-573 (2002).
[4]
RC NUCLEOTIDE SEQUENCE.
RA STRAIN=C57BL/6J; TISSUE=Aorta and vein;
```

RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,  
 RA Kono H., Akiyama J., Nishi K., Kitanai T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kaishiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Iwawa M., Ohara E., Matshiki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsumura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kita A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [6]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;  
 RA Adachi J., Aizawa K., Akimura T., Arikawa T., Bono H., Carninci P.,  
 RA Fukuda S., Furuno M., Hiramoto K., Hiraoka T., Hirozane T.,  
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Kanakawa T.,  
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,  
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohnato N., Okazaki Y.,  
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,  
 RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeeda Y., Tanaka T.,  
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK040971; BAC30765.1; -; mRNA.  
 DR MGI; MGI:1095396; Coll6a1.  
 DR GO; GO:0005737; C:cytoplasm; IEA.  
 DR GO; GO:0005578; C:extracellular matrix (sensu Metazoa); IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0007155; P:cell adhesion; IEA.  
 DR GO; GO:0006817; P:phosphate transport; IEA.  
 DR InterPro; IPR008161; C1g\_helix.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR003129; laminin\_G\_TSP\_N.  
 DR InterPro; IPR009765; Pericardin\_rpt.  
 DR Pfam; PF01391; Collagen; 13.  
 DR Pfam; PF07054; Pericardin\_rpt; 2.  
 DR ProDom; PD000007; C1g\_helix; 3.  
 DR SMART; SM00210; TSPN; 1.  
 KW Collagen; Extracellular matrix; Structural protein.  
 SQ SEQUENCE 1580 AA; 155805 MW; 3D688B9716A09050 CRC64;

Query Match 81.6%; Score 777; DB 2; Length 1580;  
 Best Local Similarity 41.5%; Pred. No. 2.3e-38;  
 Matches 152; Conservative 4; Mismatches 4; Indels 206; Gaps 2;

QY 1 MGPPGFKGTGHPGLPQPKDCCGKPPGSGTGRPAEFGPQAGMPGPPGHPGPPG 60  
 DB 1218 MGPPGFKGTGHPGLPQPKDCCGKPPGSGTGRPAEFGPQAGMPGPPGHPGPPG 1274  
 QY 61 PGQPGPAGISAVGLKGRGATGGERGLAGLPQ----- 92  
 DB 1275 PGQPGPAGISAVGLKGRGATGGERGLAGLPQ----- 1334  
 QY 93 ----- 92  
 DB 1335 GLVGPGPAGDPPGPAQKGQAGKSGRGMPPGPGKSGMGPPIGPPGPAEGRHGPSRPA 1394  
 QY 93 ----- 92  
 DB 1395 GNGPLPGLPGSGMDMNVYDDIKRFIRQEIITLFDERMAYYTSMQFMEVAAAPGRPP 1454

QY 93 -----RPPG 97  
 DB 1455 GKDGARPGAPGSPGLPQIGREGRGQLPGMRGLPGTKGKGDIVGIAGENGLGPPG 1514  
 QY 98 PGPPGKMGKATGPMGQGGIPIGPPRPMQPGKAGHCNPSDCGAMPMEQYPPMT 157  
 DB 1515 PGPPGKMGKATGPMGQGGIPIGPPRPMQPGKAGHCNPSDCGAMPMEQYPPMTS 1574  
 QY 158 MKGPPG 163  
 DB 1575 MKGPPG 1580

RESULT 6  
 Q7IRG9\_HUMAN PRELIMINARY; PRT; 218 AA.  
 AC Q7IRG9;  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE P1572.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
 OC Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.O., Qin W.X., Zhao X.T.,  
 RA Wan D.F., Gu J.R.;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF70368; AA015204.1; -; mRNA.  
 DR Ensemble; ENSG0000084636; Homo sapiens.  
 DR GO; GO:0005737; C:cytoplasm; IEA.  
 DR GO; GO:0006817; P:phosphate transport; IEA.  
 DR InterPro; IPR008161; C1g\_helix.  
 DR InterPro; IPR008160; Collagen.  
 DR Pfam; PF01391; Collagen; 3.  
 DR ProDom; PD000007; C1g\_helix; 1.  
 KW Collagen.  
 SQ SEQUENCE 218 AA; 21734 MW; 3A3C5B9194196715 CRC64;

Query Match 65.1%; Score 620; DB 2; Length 218;  
 Best Local Similarity 58.0%; Pred. No. 9.3e-30;  
 Matches 119; Conservative 7; Mismatches 37; Indels 42; Gaps 3;

QY 1 MGPPGFKGTGHPGLPQPKDCCGKPPGSGT----- 32  
 DB 14 VGPPGAGGERGHPGAGPSPGSLPGVPSMGDMNVYDEIKRFIRQEIITKFDERMAYYT 73  
 QY 33 -----RPAEFGPQAGMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 80  
 DB 74 SRMPFMEVAAAPGRGPPGKDGAPRPAAGSPGLPGQIGREGRLPGVRLPGTKG 133  
 QY 81 TGE--RGLAGLPQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 138  
 DB 134 KSDIGIGIAGENCNLPBPBPQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 193  
 QY 139 PSDCFGAMPMEQYPPMTKMGPPG 163  
 DB 194 PSDCFGAMPMEQYPPMTKMGPPG 218

RESULT 7  
 Q9CZS2\_MOUSE PRELIMINARY; PRT; 150 AA.  
 AC Q9CZS2;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Mus musculus 11 days embryo whole body cDNA, RIKEN full-length  
 DE enriched library, clone:2700007F12 product:COLLAGEN ALPHA 1 (XVI) CHAIN  
 DE homolog.

GN Name=Coll16a1;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Murinae; Mus.  
 NCBI\_TaxId=10090;  
 RN (1)  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6876(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Meth. Enzymol. 303:19-44(1999).  
 RN (2)  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H. A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gasterland T., Giesi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 RA Schriml L. M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M. F.,  
 RA Brownstein M. J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Guernicich S., Hill D., Hofmann M., Hume D. A., Kamuya M., Lee N. H.,  
 RA Lyons P., Marchionni L., Mashima U., Mazzarelli J., Momberts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K. H., Weltz C., Whitaker C., Wilming L.,  
 RA Mynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN (3)  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RA The PANTOM Consortium.  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN (4)  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN (5)  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Kono H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishize T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujitake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanishi M.,  
 RA Yonekura Y., Ishikawa T., Ozawa K., Tanaka T., Matsuyama S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 RN (6)  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Whole body;  
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,  
 RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,

RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,  
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,  
 RA Kawai J., Kojima Y., Kono H., Kouda M., Koya S., Kurihara C.,  
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,  
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,  
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,  
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,  
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
 RA Muramatsu M., Hayashizaki Y.;  
 RL Submitted (Jul-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK012212; BAB28100.1; -; mRNA.  
 DR MGI: MGI:1095396; Coll16a1.  
 DR GO: GO:0005737; Cytoplasm; IEA.  
 DR GO: GO:0005515; Protein binding; IEA.  
 DR GO: GO:0005198; Intracellular molecule activity; IEA.  
 DR GO: GO:0007155; P:cell adhesion; IEA.  
 DR GO: GO:0006817; P:phosphatase transport; IEA.  
 DR InterPro: IPR008161; C1g helix.  
 DR InterPro: IPR008160; Collagen.  
 DR Pfam: PF01391; Collagen. 2.  
 DR Prodom: PD000007; C1g\_helix; 1.  
 KW Collagen.  
 SQ SEQUENCE 150 AA; 14823 MW; D956EF9160987FC8 CRC64;  
 Query Match 58.6%; Score 558; DB 2; Length 150;  
 Best Local Similarity 67.5%; Pred. No. 3.3e-26;  
 Matches 102; Conservative 8; Mismatches 23; Indels 18; Gaps 2;  
 QY 13 PGLPGRKDPGKGRPGSGTGRPGAGEPGAMQGRPGPHVGPGRPGQPGPAGISAV 72  
 DB 18 PGRPGRPGKDGAPGRGA---PGRPOLRQIGRBRGRGLPGRKGLPGRGRGKGGDGV--- 71  
 QY 73 GLKGRGATGERGLAGLPQGRPPGRPGPGYGRKMGATGPMQCGIPGIRPGPMQGR 132  
 DB 72 -----GIAGENGLPGRPPQGRPGYGRKMGATGPMQCGIPGIRPGPMQGR 119  
 QY 133 KAGHCNPSDCRGMPEQOYPPMKTKMGKPGF 163  
 DB 120 KAGHCNPSDCRGMPEQOYPPMKSKMGKPGF 150  
 RESULT 8  
 QAS104\_7ETNG PRELIMINARY; PRT; 925 AA.  
 ID QAS104\_7ETNG PRELIMINARY; PRT; 925 AA.  
 AC QAS104;  
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
 DE Chromosome 21 SCAR14577, whole genome shotgun sequence.  
 DE (Fragment).  
 ORNames=GSTENG00017515001;  
 OS Tetraodon nigroviridis (Green puffer).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthopterygii; Acanthopterygii; Perciformes; Tetraodontiformes;  
 OC Tetraodontidae; Tetraodontidae; Tetraodon.  
 NCBI\_TaxId=99883;  
 RN (1)  
 RP NUCLEOTIDE SEQUENCE.  
 RA Jallion O., Aury J.M., Brunet F., Petit J.L., Strange-Thomann N.,  
 RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,  
 RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,  
 RA David C., Salanoubat M., Levy M., Boudet N., Castellano S.,  
 RA Anthouard V., Jubin C., Castellani V., Katinka M., Vacherie B.,  
 RA Biemont C., Skali Z., Catellico L., Poullain J., De Bernardis V.,  
 RA Cruaud C., Duprat S., Brotier P., Coulanceau J.P., Gouzy J.,  
 RA Parra G., Vardier G., Chappe C., McKernan K.J., McEwan P., Bosak S.,  
 RA Kellis M., Wolff J.N., Guigo R., Zody M.C., Nesirov J.,  
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,  
 RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,  
 RA Winkler P., Lander E.S., Weissbach J., Roest Crolius H.;  
 RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals  
 the early vertebrate proto-karyotype.";

RL Nature 431:946-957(2004).  
 RN (2)  
 RP NUCLEOTIDE SEQUENCE.  
 RG Genoscope: Whitehead Institute Centre for Genome Research;  
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
 CC -1- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL: CA601014577; CAF99438.1; -; Genomic\_DNA.  
 DR InterPro: IPR008161; C1g\_helix.  
 DR InterPro: IPR008160; C1g\_helix.  
 DR Pfam: PF01391; Collagen; 5.  
 DR ProDom: PD000007; C1g\_helix; 4.  
 DR Collagen.  
 KM NON TER  
 FT  
 SQ SEQUENCE 925 AA; 93961 MW; 4E184A0B1723BF26 CRC64;  
 Query Match 42.8%; Score 407; DB 2; Length 925;  
 Best Local Similarity 36.3%; Pred. No. 1,3e-16;  
 Matches 89; Conservative 13; Mismatches 57; Indels 86; Gaps 5;  
 QY 2 GPPGFGKTKHPLPGKKGCGKPGPGSTGRPAEGEP--GAMGPGGRGPPGHVPPG 59  
 DB 675 GSPGQKGPSPGQKSGPGRGPPGGGYSKKAQPMVGPAGPGRGERSPSESGIPG 734  
 QY 60 PPGGPPAGISAV----- 72  
 DB 735 SPFLPSPGNDVYVNDIEIKFIRQVYIKVFDERMAVYMRQRPVMAASPGRPAPGK 794  
 QY 73 -----GLKGRDGAATGERSIAGLPQGPDPGPPG 101  
 DB 795 DGAPGSPGLPGTGRPGHLPGRQGLPGSQGMPGPGKDGKDGKIGAGDTPPGAGPVP 854  
 QY 102 --PGYGMGATGPMGQGGIGIPIPGPPPMQPGKAGCNSDPCGAMPMQQTPPM-KTM 158  
 DB 855 GAGCYGMGAPGAPVGGQGGVGLPGPPGHPGSGKSIDRCNPDGDC---VSHQVDYSPKGPVS 911  
 QY 159 KGPFG 163  
 DB 912 KGPWG 916  
 RESULT 9  
 Q8BNS7 MOUSE  
 ID Q8BNS7\_MOUSE PRELIMINARY; PRT: 799 AA.  
 AC Q8BNS7;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Mus musculus 10 days neonate cortex cDNA, RIKEN full-length enriched  
 DE library, clone: A830093P02 product: procollagen, type IV, alpha 5, full  
 DE insert sequence. (Fragment).  
 DE Name=Col4a5;  
 GN Mus musculus (Mouse).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Murinae; Mus.  
 OC NCB1\_TaxID=10090;  
 RN (1)  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Cortex;  
 RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.,  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Mech. Enzymol. 303:13-44(1999).  
 RN (2)  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Cortex;  
 RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi U., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Pleischmann W., Gaasterland T., Giesi C., King B., Kochwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Guentrich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,  
 RA Wyshak-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN (3)  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Cortex;  
 RA The PANTOM Consortium.  
 RA "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN (4)  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Cortex;  
 RC MEDLINE=2049374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subcloning of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN (5)  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Cortex;  
 RC MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Itch M.,  
 RA Kono H., Akiyama J., Nishi K., Kiteunai T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 RN (6)  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=C57BL/6J; TISSUE=Cortex;  
 RA Adachi U., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoaka T., Hirazane T.,  
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,  
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata N., Nakamura M.,  
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohnishi N., Okazaki Y.,  
 RA Saito K., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
 RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,  
 RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,  
 RA Tomaru A., Tova T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK080682; BNC37980.1; -; mRNA.  
 DR HSSP: P02462; IL11.  
 DR SMR: Q8BNS7; 572-799.  
 DR MGI: MGI:88456; Col4a5.  
 DR GO: GO:0005604; C:basement membrane; IDA.  
 DR GO: GO:0005615; C:extracellular space; TMS.  
 DR InterPro: IPR008161; C1g\_helix.  
 DR InterPro: IPR008160; Collagen.  
 DR Pfam: PF01413; C4; 2.  
 DR Pfam: PF01391; Collagen; 9.

DR ProDom; PD000007; C1g\_helix; 2.  
DR ProDom; PD003923; Procollagn4\_C; 2.  
DR SMART; SM00111; C4; 2.  
KW Collagen.  
FT NON\_TER  
SQ SEQUENCE 799 AA; 77889 MW; C517CFACF15706DC CRC64;

Query Match 41.1%; Score 391; DB 2; Length 799;  
Best Local Similarity 49.7%; Pred. No. 1e-15;  
Matches 82; Conservative 13; Mismatches 52; Indels 18; Gaps 5;

QY 2 GPPGFGKGTGHPGLPCKDCKPGRPGSTGRBAGEGAPQGRPPGHVGPSP 61  
DB 105 GPPGSLGGQSPGLPCKGNPGLPGLPGLPGLPGLPGLPGLPGLPGLP 161  
QY 62 GPPGAGISAVGLKGRGATGERGLA--GLPGQPPGPPGQPPGY-GKMGATGPMQOQGI 118  
DB 162 GPPGQPG--SPGLPGQKSKGKRGVSGITLPGLPGRKSGPLGPIYGNIGKSGVGTGL 219  
QY 119 PGIPGPPGPMGQPGKAGHCNPSPDCFGAMPMEQYPPMTKMPFG 163  
DB 220 PGLPGNPGAKGQGLPG-----PFGTGLPGPKMNGPPG 254

RESULT 10  
Q9ESQ2 MOUSE PRELIMINARY; PRT; 1691 AA.

ID Q9ESQ2 MOUSE PRELIMINARY; PRT; 1691 AA.  
AC Q9ESQ2;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Type IV collagen alpha 5 chain.  
GN Name=C01445;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=20356494; PubMed=10965041;  
RA Saiko K., Naito I., Seki T., Ohashi T., Kimura E., Momota R.,  
RA Kishino Y., Sado Y., Yoshioka H., Ninomiya Y.,  
RA "Differential Expression of Mouse a5(IV) and a6(IV) Collagen Genes in  
RT Epithelial Basement Membranes.";  
RL J. Biochem. 128:427-434(2000).  
DR EMBL; AB041350; BAB13673.1; -; mRNA.  
DR HSSP; P02462; 1L11.  
DR SMR; Q9ESQ2; 1464-1691.  
DR Ensemble; ENSMUSG00000031274; Mus musculus.  
DR MGI; MGI:88456; C01445.  
DR GO; GO:0005604; C:basement membrane; IDA.  
DR GO; GO:0005615; C:extracellular space; TAS.  
DR InterPro; IPR008161; C1g\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 24.  
DR ProDom; PD000007; C1g\_helix; 3.  
DR ProDom; PD003923; Procollagn4\_C; 2.  
DR SMART; SM00111; C4; 2.  
KW Collagen.  
SQ SEQUENCE 1691 AA; 161823 MW; 81340DF1792208FA CRC64;

Query Match 41.1%; Score 391; DB 2; Length 1691;  
Best Local Similarity 49.7%; Pred. No. 1.9e-15;  
Matches 82; Conservative 13; Mismatches 52; Indels 18; Gaps 5;

QY 2 GPPGFGKGTGHPGLPCKDCKPGRPGSTGRBAGEGAPQGRPPGHVGPSP 61  
DB 991 GPPGSLGGQSPGLPCKGNPGLPGLPGLPGLPGLPGLPGLPGLPGLP 1047  
QY 62 GPPGAGISAVGLKGRGATGERGLA--GLPGQPPGPPGQPPGY-GKMGATGPMQOQGI 118

DB 1048 GPPGQPG--SPGLPGQKSKGKRGVSGITLPGLPGRKSGPLGPIYGNIGKSGVGTGL 1105  
QY 119 PGIPGPPGPMGQPGKAGHCNPSPDCFGAMPMEQYPPMTKMPFG 163  
DB 1106 PGLPGNPGAKGQGLPG-----PFGTGLPGPKMNGPPG 1140

RESULT 11  
Q632W6 MOUSE PRELIMINARY; PRT; 1691 AA.

ID Q632W6;  
AC Q632W6;  
DT 25-OCT-2004 (TREMBLrel. 28, Created)  
DT 25-OCT-2004 (TREMBLrel. 28, Last sequence update)  
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)  
DE Procollagen, type IV, alpha 5.  
GN Name=C01445;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stjepicono L., Soares M.B., Bonaldo M.F., Caavaant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huijck S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahney J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;  
RG NIH MGC Project;  
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC082788; AAH82788.1; -; mRNA.  
DR SMR; Q632W6; 1464-1691.  
DR Ensemble; ENSMUSG00000031274; Mus musculus.  
DR MGI; MGI:88456; C01445.  
DR GO; GO:0005581; C:collagen; IEA.  
DR GO; GO:0005737; C:cycloplassm; IEA.  
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
DR GO; GO:0006817; P:phosphate transport; IEA.  
DR InterPro; IPR008161; C1g\_helix.  
DR InterPro; IPR008160; Collagen.  
DR InterPro; IPR009765; Pericardial\_rpt.  
DR InterPro; IPR001442; Procollagn4\_C.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 21.  
DR Pfam; PF07054; Pericardial\_rpt; 2.  
DR ProDom; PD000007; C1g\_helix; 3.  
DR ProDom; PD003923; Procollagn4\_C; 2.  
DR SMART; SM00111; C4; 2.  
KW Collagen.  
SQ SEQUENCE 1691 AA; 161807 MW; B5CD42658128DD07 CRC64;

Query Match 41.1%; Score 391; DB 2; Length 1691;



Best Local Similarity 49.7%; Pred. No. 1.9e-15;  
Matches 82; Conservative 13; Mismatches 52; Indels 18; Gaps 5;  
QY 2 GPPFGKXGKHPGLPGPKGDCGKRGPRGSGRGABSEBPAMRGQRGPRGHVGRPPG 61  
DB 991 GQGLSGQPSPLPGPKGPGGLPGPPGLTGPPKKN---IDMGPPGQGVDPGPP 1047  
QY 62 GQGPAGISAVGKSGRGATGREGLA--GLPGQPGPPGQPPGY--GKMGATGPMGQGI 118  
DB 1048 GFGPGGCG--SPGLPGQKSGKSGEPGVSIGLPGLPGRPGEGELGCPGNPGLKSGVETGL 1105  
QY 119 PGIPGPPGPGQPGKAGHCNPSCFCGAMPBQGYPPMKTGKGGPG 163  
DB 1106 PGIPGPPGKAGQPGGLPG-----FPGTGLPGPKGNMGPPG 1140  
RESULT 12  
CO4A5\_HUMAN STANDARD; PRT; 1685 AA.  
ID CO4A5\_HUMAN P29400; Q16006; Q16126; O6LDB4;  
AC P29400; Q16006; Q16126; O6LDB4;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 10-MAY-2005 (Rel. 47, Last annotation update)  
DE Collagen alpha 5 (IV) chain precursor.  
GN Name=COL4A5;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=94165049; PubMed=8120014;  
RA Zhou J., Leinonen A., Tryggvason K.;  
RT "Structure of the human type IV collagen COL4A5 gene";  
RL J. Biol. Chem. 269:6608-6614(1994).  
RN [2]  
RP NUCLEOTIDE SEQUENCE OF 1-910, AND VARIANT AS CYS-521.  
RC TISSUE=Kidney;  
RX MEDLINE=92316923; PubMed=1352287;  
RA Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;  
RT "Complete amino acid sequence of the human alpha 5 (IV) collagen chain  
and identification of a single-base mutation in exon 23 converting  
glycine 521 in the collagenous domain to cysteine in an Alport  
syndrome patient";  
RL J. Biol. Chem. 267:12475-12481(1992).  
RN [3]  
RP NUCLEOTIDE SEQUENCE OF 85-1685.  
RC TISSUE=Placenta;  
RX MEDLINE=90337990; PubMed=2380186;  
RA Pihlajaniemi T., Pohjolainen E.R., Myers J.C.;  
RT "Complete primary structure of the triple-helical region and the  
carboxyl-terminal domain of a new type IV collagen chain, alpha  
5(IV).";  
RL J. Biol. Chem. 265:13758-13766(1990).  
RN [4]  
RP NUCLEOTIDE SEQUENCE OF 924-1685.  
RX MEDLINE=91169491; PubMed=2004755;  
RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;  
RT "Characterization of the 3' half of the human type IV collagen alpha 5  
gene that is affected in the Alport syndrome";  
RL Genomics 9:1-9(1991).  
RN [5]  
RP NUCLEOTIDE SEQUENCE OF 914-1685.  
RX MEDLINE=90160375; PubMed=1689491;  
RA Hostikka S.L., Eddy R.L., Byers M.G., Hoeyhtyae M., Shows T.B.,  
RA Tryggvason K.;  
RT "Identification of a distinct type IV collagen alpha chain with  
restricted kidney distribution and assignment of its gene to the locus  
of X chromosome-linked Alport syndrome";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).  
RN [6]  
RP NUCLEOTIDE SEQUENCE OF 1442-1471.

RX MEDLINE=90252791; PubMed=2339699;  
RA Myers J.C., Jones T.A., Pohjolainen E.R., Kadri A.S., Goddard A.D.,  
RA Sheer D., Solomon E., Pihlajaniemi T.;  
RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene  
to the region of the X chromosome containing the Alport syndrome  
locus";  
RL Am. J. Hum. Genet. 46:1024-1033(1990).  
RN [7]  
RP NUCLEOTIDE SEQUENCE OF 1-20.  
RA Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J.,  
RA Marynen P.;  
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
RN [8]  
RP NUCLEOTIDE SEQUENCE OF 1258-1270 (ISOFORM 2).  
RX MEDLINE=94133540; PubMed=8301933;  
RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H.,  
RA Cassiman J.-J., Marynen P.;  
RT "Differential splicing of COL4A5 mRNA in kidney and white blood cells:  
a complex mutation in the COL4A5 gene of an Alport patient deletes the  
NC1 domain";  
RL Kidney Int. 44:1316-1321(1993).  
RN [9]  
RP NUCLEOTIDE SEQUENCE OF 1589-1598 AND 1677-1685, AND VARIANTS AS  
RP 1597-TYR--THR-1685 DEL AND 1679-GLU--THR-1685 DEL.  
RX PubMed=7853788;  
RA Nakazato H., Hattori S., Ushijima T., Matsura T., Kotabashi Y.,  
RA Takada T., Yoshioke K., Endo F., Matsuda I.;  
RT "Mutations in the COL4A5 gene in Alport syndrome: a possible mutation  
in primordial germ cells";  
RL Kidney Int. 46:1307-1314(1994).  
RN [10]  
RP REVIEW ON VARIANTS.  
RX MEDLINE=9733662; PubMed=9195222;  
RA DOI=10.1002/(SICI)1098-1004(1997)9:6<477::AID-HUMU13.3.CO;2-H;  
RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;  
RT "The clinical spectrum of type IV collagen mutations";  
RL Hum. Mutat. 9:477-499(1997).  
RN [11]  
RP VARIANT AS SER-1564.  
RX MEDLINE=91169492; PubMed=1672282;  
RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L.,  
RA Tryggvason K.;  
RT "Single base mutation in alpha 5(IV) collagen chain gene converting a  
RT conserved cysteine to serine in Alport syndrome";  
RL Genomics 9:10-18(1991).  
RN [12]  
RP VARIANT AS ARG-325.  
RX MEDLINE=92303559; PubMed=1376965;  
RA Knebelmann B., Deschenes G., Gros F., Hore M.-C., Gruenfeld J.-P.,  
RA Tryggvason K., Gubler M.-C., Aulagnac C.;  
RT "Substitution of arginine for glycine 325 in the collagen alpha 5 (IV)  
chain associated with X-linked Alport syndrome: characterization of  
RT the mutation by direct sequencing of PCR-amplified lymphoblast cDNA  
fragments";  
RL Am. J. Hum. Genet. 51:135-142(1992).  
RN [13]  
RP VARIANT AS GLU-325.  
RX MEDLINE=93244772; PubMed=1363780;  
RA Renieri A., Serti M., Myers J.C., Pihlajaniemi T., Maseella L.,  
RA Rizzoni G.F., de Marchi M.;  
RT "De novo mutation in the COL4A5 gene converting glycine 325 to  
RT glutamic acid in Alport syndrome";  
RL Hum. Mol. Genet. 1:127-129(1992).  
RN [14]  
RP VARIANTS AS THR-1517, SER-1538 AND GLN-1563.  
RX MEDLINE=94010948; PubMed=8406498;  
RA Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J.,  
RA Tryggvason K., Haggama-Schouten W.A.G., Roodvoets A.P., Rascher W.,  
RA van Oost B.A., Smeets H.J.M.;  
RT "Identification of four novel mutations in the COL4A5 gene of patients  
RT with Alport syndrome";  
RL Genomics 17:485-489(1993).  
RN [15]

RP VARIANTS AS GLU-400; VAL-406; VAL-638; ALA-638; ARG-653; ARG-796;  
 RP ARG-869; ARG-872 AND CYS-1241.  
 RX MEDLINE=95322976; PubMed=7599631;  
 RA Boyle E., Flinter F., Zhou J., Trygvason K., Bobrow M., Harris A.;  
 RT "Detection of 12 novel mutations in the collagenous domain of the  
 RL COL4A5 gene in Alport syndrome patients."; Hum. Mutat. 5:197-204(1995).  
 RN [16]  
 RP VARIANT AS ARG-1649.  
 RX MEDLINE=966213750; PubMed=8651292;  
 RA Barber D.P., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,  
 RT "A mutation causing Alport syndrome with tardive hearing loss is  
 RL common in the western United States."; Am. J. Hum. Genet. 58:1157-1165(1996).  
 RN [17]  
 RP VARIANTS AS  
 RX MEDLINE=96213754; PubMed=8651296;  
 RA Renieri A., Bruttini M., Galli L., Zanelli P., Neri T.M., Rossetti S.,  
 RX Turco A.E., Heiskari N., Zhou J., Gushima R., Massella L., Banfi G.,  
 RA Scialari F., Sessa A., Rizzoni G.F., Trygvason K., Pignatti P.F.,  
 RA Savi M., Ballabio A., de Marchi M.;  
 RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51  
 RL exons of the COL4A5 gene."; Am. J. Hum. Genet. 58:1192-1204(1996).  
 RN [18]  
 RP VARIANTS AS, AND VARIANTS ASP-430; SER-444; SER-619; ASN-664 AND  
 RP MET-1428.  
 RX MEDLINE=97094179; PubMed=8940267;  
 RA Knebelmann B., Breillat C., Forestier L., Arrondel C., Jaccassier D.,  
 RA Glazas I., Drouot L., Deschenes G., Gruenfeld J.-P., Broyer M.,  
 RA Gubler M.-C., Antignac C.;  
 RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport  
 RL syndrome."; Am. J. Hum. Genet. 59:1221-1232(1996).  
 RN [19]  
 RP VARIANT AS ASP-1498.  
 RX MEDLINE=96233932; PubMed=8829632;  
 RL DOI=10.1002/(SICI)1098-1004(1996)7:2<149::AID-HUMU9>3.3.CO;2-A;  
 RX Tverskaya S., Bobrykina V., Tsalykova F., Ignatova M.,  
 RA Krasnopolskaya X., Bygratov O.;  
 RT "Substitution of Al498D in noncollagen domain of  $\alpha 5(IV)$  collagen chain  
 RL associated with adult-onset X-linked Alport syndrome."; Hum. Mutat. 7:149-150(1996).  
 RN [20]  
 RP VARIANT AS GLN-1677.  
 RX MEDLINE=97295089; PubMed=9150741; DOI=10.1007/s004390050429;  
 RA Barber D.F., Denison J.C., Atkin C.L., Gregory M.C.;  
 RT "Common ancestry of three Ashkenazi-American families with Alport  
 RL syndrome and COL4A5 R1677Q."; Hum. Genet. 99:681-684(1997).  
 RN [21]  
 RP VARIANTS AS ARG-174; ARG-177; ARG-325; CYS-1410; TRP-1421; THR-1517  
 RP AND ASP-1596.  
 RX MEDLINE=98112435; PubMed=9452056;  
 RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,  
 RA Pignatti G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,  
 RA Trivelli A., Pinciaroli A.R., Ragazzo M., Rizzoni G.F., de Marchi M.;  
 RT "Missense mutations in the COL4A5 gene in patients with X-linked  
 RL Alport syndrome."; Hum. Mutat. Suppl. 1:S106-S109(1998).  
 RN [22]  
 RP VARIANTS AS VAL-420; 456-PRO-PRO-458 DEL; ASP-573; ASP-624; ASP-635;  
 RP 802-GLY-PRO-807 DEL; ARG-869; CYS-941; SER-1030; SER-1066; ASP-1143;  
 RX ARG-1196; GLU-1261; SER-1357 AND ARG-1649.  
 RP MEDLINE=99063529; PubMed=9848783;  
 RA Martin P., Heiskari N., Zhou J., Leinonen A., Tanelius T., Hertz J.M.,  
 RA Barker D.F., Gregory M.C., Atkin C.L., Stykardottir U., Neumann H.,  
 RA Sprigade J., Shows T.B., Petersen E., Trygvason K.;  
 RT "High mutation detection rate in the COL4A5 collagen gene in suspected  
 RL Alport syndrome using PCR and direct DNA sequencing."; J. Am. Soc. Nephrol. 9:2291-2301(1998).

Query Match 41.0%; Score 390; DB 1; Length 1685;  
 Best Local Similarity 43.2%; Pred. No. 2.2e-15;  
 Matches 80; Conservative 11; Mismatches 40; Indels 54; Gaps 4;  
 QY 2 GPPGKGTGTPG-----LPPKGDCKPAPPSTGTGPGAGEPGAMGPPGPPGHV 55  
 DB 147 GPPGIPGKKGKBPSTIMSLPFGPKNPGYPPPGIGLPGTGTGPGIGP-----PPPLM 203  
 QY 56 GPPPPPGPPAGISAVGLKDRGATGERTGLP----- 90  
 DB 204 GPPGPPGLPGPKGNMGNLFGPKPKBKEGGLGPPGPPGQISEQRPIDVEFGKDDGLP 263  
 QY 91 ---GPPGPPGPPG-----YKMGATGPMGQGGIIGIIPGPPMGQ 130  
 DB 264 GDRGPPGPPGIRGPPGPKGKEXGKQEPGPKRGPKGDBNGQGIPLGPDYFGE 323  
 QY 131 PGKAG 135  
 DB 324 PGKDG 328  
 RESULT 13  
 Q9NIB7 HUMAN PRELIMINARY; PRT; 1685 AA.  
 ID Q9NIB7 Q72700; 01-OCT-2000 (TrEMBLrel. 15, Created)  
 AC Q9NIB7 Q72700; 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
 DE Collagen, type IV, alpha 5 (Alport syndrome).  
 GN Name=COL4A5; ORFNames=RP6-24A23.5-001;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae;  
 OC Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Cobley V.;  
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.  
 RP [2]  
 RN NUCLEOTIDE SEQUENCE.  
 RA Bird C.;  
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL035425; CAB90289.2; -; Genomic DNA.  
 DR EMBL; AL034369; CAA22267.2; -; Genomic DNA.  
 DR EMBL; AL031622; CA143038.1; -; Genomic DNA.  
 DR EMBL; AL034629; CAB90289.2; JOINED; Genomic DNA.  
 DR EMBL; AL031622; CAA22267.2; JOINED; Genomic DNA.  
 DR EMBL; AL034625; CAA22267.2; JOINED; Genomic DNA.  
 DR EMBL; AL034369; CA143038.1; JOINED; Genomic DNA.  
 DR EMBL; AL035425; CA143038.1; JOINED; Genomic DNA.  
 DR SMR; Q9NIB7; 1458-1685.  
 DR Ensemble; ENSG00000188153; Homo sapiens.  
 DR GO; GO:0005581; C:collagen; IEA.  
 DR GO; GO:0005737; C:cytoplasm; IEA.  
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.  
 DR GO; GO:0006817; F:phosphate transport; IEA.  
 DR InterPro; IPR008161; C1g helix.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR001442; Procollagen4\_C.  
 DR Pfam; PF01413; C4; 2.  
 DR Pfam; PF01391; Collagen; 22.  
 DR ProDom; PD000007; C1g helix; 3.  
 DR ProDom; PD0003923; Procollagen4\_C; 2.  
 DR SMART; SM00111; C4; 2.  
 KW collagen.  
 SQ SEQUENCE 1685 AA; 161044 MW; 4450A6762F12A626 CRC64;  
 Query Match 41.0%; Score 390; DB 2; Length 1685;  
 Best Local Similarity 43.2%; Pred. No. 2.2e-15;  
 Matches 80; Conservative 11; Mismatches 40; Indels 54; Gaps 4;

QY 2 GPPGFKGTGHPG-----LPGKGDCKRPPGSGTGRPGAGEPGAMGPGRPGGHV 55  
 DB 147 GPPGIGMKGEPGSIIMSLPGKGNPGVPGPGIGQLPPTGIPGPIGP---PGPPGLM 203  
 QY 56 GPPGPPGPGPGNAGISAVGLKGDGATGERGLAP----- 90  
 DB 204 GPPGPPGGLPBPKNMGNLFGKPGKEGQGLGPPGPGQISQKPRIDVFGKGDGLP 263  
 QY 91 ---GPPGPPGPPGPPG-----YKMGATGPMGQGGIPGIPGPPGPMGQ 130  
 DB 264 GDDGPPGPPGPIRPPRPPGSGKEKGEQGRPGKRGKRGKXGNGGPGIPLPDPGPIRGE 323  
 QY 131 PGKAG 135  
 DB 324 PGKAG 328

RESULT 14  
 COH1\_CHICK STANDARD; PRT; 1146 AA.  
 ID COH1\_CHICK  
 AC 090584;  
 DT 29-MAR-2004 (Rel. 43, Last sequence update)  
 DT 29-MAR-2004 (Rel. 43, Last sequence update)  
 DT 10-MAY-2005 (Rel. 47, Last annotation update)  
 DE Collagen alpha 1(XVII) chain (Bullous pemphigoid antigen 2) (180 kDa  
 DE bullous pemphigoid antigen 2) (Fragment).  
 GN Name=COL17A1; Synonyms=BP180, BPAG2;  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE, AND TISSUE SPECIFICITY.  
 RC TISSUE=Cornea;  
 RX MEDLINE=91142213; PubMed=1705041;  
 RA Marchant J.K., Linsmayer T.F., Gordon M.K.;  
 RT "CDNA analysis predicts a cornea-specific collagen.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:1560-1564(1991).  
 CC -1- FUNCTION: May play a role in the integrity of hemidesmosome and  
 CC the attachment of basal keratinocytes to the underlying basement  
 CC membrane (By similarity).  
 CC -1- SUBUNIT: Homotrimer of alpha 1(XVII) chains (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (By similarity).  
 CC -1- TISSUE SPECIFICITY: Cornea specific.  
 CC -1- PTM: The intracellular/endo domain is disulfide-linked (By  
 CC similarity).  
 CC -1- PTM: Prolines at the third position of the tripeptide repeating  
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
 CC -1- PTM: Undergoes proteolytic cleavage by furin-like protease to  
 CC yield a 120 kDa soluble form that corresponds to the ectodomain  
 CC (By similarity).  
 CC -----  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use as long as its content is in no way modified and this statement is not  
 CC removed.  
 CC -----  
 CC EMBL; M60172; AAA48703.1; -; mRNA.  
 DR PIR; S16501; A38587.  
 DR Ensemble; ENSGALG00000008323; Gallus gallus.  
 DR InterPro; IPR008161; C1g\_helix.  
 DR InterPro; IPR008160; Collagen.  
 DR Pfam; PF01391; Collagen; 6.  
 DR Prodom; PD000007; C1g\_helix; 3.  
 KW Collagen; Extracellular matrix; Glycoprotein; Hydroxylation; Repeat;  
 KW Signal-anchor; Structural protein; Transmembrane.  
 FT TOPO\_DOM 41 52 Cytoplasmic (Potential).  
 FT TRANSMEM 53 74 Signal-anchor for type II membrane  
 FT TOPO\_DOM 75 1146 protein (Potential).  
 FT REGION 1 149 Extracellular region (Potential).  
 FT Nonhelical region (NC16).

FT REGION 150 1101 Triple-helical region.  
 FT REGION 1102 1146 Nonhelical region (NC1).  
 FT CARBOHYD 1041 1041 N-linked (GlcNAc...) (Potential).  
 FT NON\_TER 1 1  
 SQ SEQUENCE 1146 AA; 114814 MW; 1E0E9977D537A859 CRC64;  
 Query Match 40.7%; Score 387; DB 1; Length 1146;  
 Best Local Similarity 40.1%; Pred. No. 2,4e-15;  
 Matches 89; Conservative 10; Mismatches 57; Indels 66; Gaps 7;  
 QY 2 GPPGFKGTGHPGLPBPCKDCKRPPGSGTGRPGAGEPGAMGPPG-----QGRP 49  
 DB 350 GPPGPGDGTGEBPLTGPQ---GPPGLPGNPGRGAGGEPGAPGVASAGSSTIALPGPP 406  
 QY 50 GPPGHPGPPGPPGPPGPPGAGISAVGLKGDGATGERG-----LAG 88  
 DB 407 GPPGPIGPTGPPGVPPVG---PAGLPQGGPGEKGSANVAVETITKTEVSSIASQMLSD 464  
 QY 89 LPQPPGPPGPPGPPGPPGKMGATGPMGQGGIP-----GIPGPPG 127  
 DB 465 LQGRAGPPGPPGPPGSSVQGLPGRPRPGLPGRSGPRGRGSSVSTSTFVSGPPGPPG 524  
 QY 128 MGQPGKAGHCNPSDCRGAMPMEQQYPMK-----TMKPPG 163  
 DB 525 PGKPGDGPSPRG-FTGEPFEGGLPFGSSHGGLVTVMQPPG 565

RESULT 15  
 COH1\_CHICK STANDARD; PRT; 1453 AA.  
 ID COH1\_CHICK  
 AC P02457;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 10-MAY-2005 (Rel. 47, Last annotation update)  
 DE Collagen alpha 1(I) chain precursor.  
 GN Name=COL1A1;  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE OF 1-153.  
 RX MEDLINE=88056316; PubMed=3678834; DOI=10.1016/0378-1119(87)90159-4;  
 RA Finer M.H., Boedtker H., Doty P.;  
 RT "Unusual DNA sequences located within the promoter region and the  
 RT first intron of the chicken pro-alpha 1(I) collagen gene.";  
 RL J. Biol. Chem. 262:13323-13332(1987).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE OF 1-144.  
 RX MEDLINE=88007542; PubMed=2820966;  
 RA Finer M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;  
 RT "Unusual DNA sequences located within the promoter region and the  
 RT first intron of the chicken pro-alpha 1(I) collagen gene.";  
 RL J. Biol. Chem. 262:13323-13332(1987).  
 RN [3]  
 RP PROTEIN SEQUENCE OF 152-1187.  
 RX MEDLINE=82231995; PubMed=7093229;  
 RA Hildeberger J.H., Corbett C., Dixit S.N., Yu W., Seyer J.M., Kang A.H.,  
 RA Gross J.;  
 RT "Amino acid sequence of chick skin collagen alpha 1(I)-C88 and the  
 RT complete primary structure of the helical portion of the chick skin  
 RT collagen alpha 1(I) chain.";  
 RL Biochemistry 21:2048-2055(1982).  
 RN [4]  
 RP PROTEIN SEQUENCE OF 1200-1205.  
 RX MEDLINE=72243016; PubMed=5047697;  
 RA Eyre D.R., Glimcher M.J.;  
 RT "Evidence for a previously undetected sequence at the carboxyterminus  
 RT of the alpha 1 chain of chicken bone collagen.";  
 RL Biochem. Biophys. Res. Commun. 48:720-726(1972).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE OF 981-1453.

MDLINE=81160715; PubMed=6927845;  
RA Fuller F., Boedtker H.;  
RT "Sequence determination and analysis of the 3' region of chicken pro-  
alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids  
including the carboxy-terminal propeptide sequences.";  
RL Biochemistry 20:996-1006(1981).  
[6]  
RP NUCLEOTIDE SEQUENCE OF 1311-1453.  
RX MEDLINE=80134546; PubMed=6987088; DOI=10.1016/0014-5793(80)80761-7;  
RA Showalter A.M., Pesciotta D.M., Eikenberry E.F., Yamamoto T.,  
RA Pastan I., Derombrughe B., Pietrek P.P., Olsen B.R.;  
RT "Nucleotide sequence of a collagen cDNA-fragment coding for the  
carboxyl end of pro alpha 1(I)-chains.";  
RL FEBS Lett. 111:61-65(1980).  
CC -!- FUNCTION: Type I collagen is a member of group I collagen  
(fibrillar forming collagen).  
CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.  
CC -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and  
bones. In bones the fibrils are mineralized with calcium  
hydroxyapatite.  
CC -!- PTM: Proline residues at the third position of the tripeptide  
repeating unit (G-X-Y) are hydroxylated in some or all of the  
chains. Pro-1153 is the only 3-hydroxypro and the only  
hydroxylated proline in position X.  
CC -!- SIMILARITY: Belongs to the fibrillar collagen family.  
CC -!- SIMILARITY: Contains 1 WWC domain.  
-----  
CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use as long as its content is in no way modified and this statement is not  
removed.  
-----  
CC EMBL: M17839; AAA48704.1; -; Genomic DNA.  
CC EMBL: M17838; AAA48704.1; JOINED; Genomic DNA.  
CC EMBL: V00401; CA23695.1; -; mRNA.  
CC EMBL: M10571; AAA48671.1; ALT\_SEQ; mRNA.  
CC EMBL: M17607; AAA48672.1; -; mRNA.  
CC PIR: A27179; A27179.  
CC PIR: I50629; I50629.  
CC InterPro: IPR008161; Clg\_helix.  
CC InterPro: IPR008160; Collagen.  
CC InterPro: IPR000885; Fib\_collagen\_C.  
CC InterPro: IPR001007; WWC\_C.  
CC Pfam: PF01410; COLFI; 1.  
CC Pfam: PF01391; Collagen; 18.  
CC Pfam: PF00093; WWC; 1.  
CC ProDom: PD000007; Clg\_helix; 2.  
CC ProDom: PD002078; Fib\_collagen\_C; 1.  
CC SMART; SM00214; COLFI; 1.  
CC SMART; SM00214; WWC; 1.  
CC PROSITE; PS01208; WWC\_1; 1.  
CC PROSITE; PS0184; WWC\_2; 1.  
CC Collagen; Direct protein sequencing; Extracellular matrix;  
KW Glycoprotein; Hydroxylation; Pyrolydine carboxylic acid; Repeat;  
KW Signal; Structural protein.  
FT SIGNAL 1 22  
FT PROPEP 23 151 N-terminal propeptide.  
FT CHAIN 152 1205 Collagen alpha 1(I) chain.  
FT PROPEP 1206 1453 C-terminal propeptide.  
FT DOMAIN 31 89 WWC.  
FT MOD\_RES 152 152 Pyrolydine carboxylic acid.  
FT MOD\_RES 160 160 Alysine (By similarity).  
FT MOD\_RES 254 254 5-hydroxylysine (By similarity).  
FT MOD\_RES 851 851 5-hydroxylysine (Potential).  
FT MOD\_RES 1081 1081 Hydroxyproline (Potential).  
FT MOD\_RES 1097 1097 5-hydroxylysine (Potential).  
FT MOD\_RES 1153 1153 3-hydroxyproline.  
FT CARBOHYD 254 254 O-linked (Gal.. ) (By similarity).  
FT CARBOHYD 1354 1354 N-linked (GlcNAc.. ) (By similarity).  
FT CONFLICT 1187 1187 F -> L (in Ref. 5).  
FT CONFLICT 1441 1441 Q -> H (in Ref. 6).  
SQ SEQUENCE 1453 AA; 137789 MW; 3BC6152134271F4D CRC64;

Query Match 40.7%; Score 387; DB 1; Length 1453;  
Best Local Similarity 46.9%; Pred. No. 2.9e-15;  
Matches 82; Conservative 12; Mismatches 59; Indels 22; Gaps 5;  
QY 2 GPP---GFGKTHGHPGLPPKDCGKPPPGSTGRPGAEGEPGAMGPPGPPGHVGP 58  
DB 858 GPPGATGPPGAAGVGPVGPVPSGNITGLPPPPACKZSKGRGRTGPPAGRGPPAGPP 917  
QY 59 GPPGQPPGAG---ISAVGLKDRGANGERGLAGLPGQ-----PGPPGQPPGEGKMG 108  
DB 918 GPPGEGKSPGADGPIGAPGTPGPGIAGQGVGLPGQRGGRGFPGLGPPSGEP--GKQG 975  
QY 109 ATGPMGQCGITGIRPPPPPMQPGKAGHCNPSDCGAMPMEQYPPMKTKMGPPG 163  
DB 976 PSGASGERGPPGMPGPPGIAGPPGEGREGAPGAEGA-----PGRDGAAGPKG 1023  
Search completed: December 1, 2005, 08:26:47  
Job time : 233 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Comugen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2005, 08:28:25 ; Search time 188 seconds  
(without alignments)  
380.951 Million cell updates/sec

Title: US-09-924-340-58

Sequence: 1 MGPPGFKGTGHPGLPGPKG.....GAMPMEQOYPMKTMGPRG 163

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 2443163 seqs, 439378781 residues

Word size : 100

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

A\_Geneseq\_21:\*

- 1: geneseqp1980s:\*
- 2: geneseqp1980s:\*
- 3: geneseqp2000s:\*
- 4: geneseqp2001s:\*
- 5: geneseqp2002s:\*
- 6: geneseqp2003as:\*
- 7: geneseqp2003bs:\*
- 8: geneseqp2004s:\*
- 9: geneseqp2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	163	100.0	163	ABR48481	ABR48481 Human Alp

#### ALIGNMENTS

RESULT 1  
ABR48481  
ID ABR48481 standard; protein; 163 AA.  
XX  
AC ABR48481;  
XX  
DT 13-JUN-2003 (first entry)  
XX  
DE Human Alpha 1 type XVI collagen.  
XX  
KW Human; GENSET; therapeutic; therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO200294864-A2.  
XX  
PD 28-NOV-2002.  
XX

PF 06-AUG-2001; 2001WO-1B001715.  
XX  
PR 25-MAY-2001; 2001US-0293574P.  
PR 15-JUN-2001; 2001US-0298698P.  
PR 29-JUN-2001; 2001US-0302277P.  
PR 13-JUL-2001; 2001US-0305456P.  
XX

PA (GEST ) GENSET.

PI Benjamin S, Tanaka H;

XX WPI; 2003-129412/12.  
DR N-PDB; ACC51088.  
XX

PT New GENSET polynucleotides and polypeptides, useful for preparing a composition for treating GENSET-related disorders and as reagents in assays to quantitatively determined levels of GENSET expression in biological samples.  
PT  
XX  
PS Claim 2; Page 451; 505pp; English.

CC The present invention relates to novel human GENSET coding sequences (ACC51060-ACC51115) and proteins (ABR48453-ABR48508). The GENSET sequences are useful for preparing a composition for treating GENSET-related disorders. They can also be used as markers for tissues in which the corresponding protein is preferentially expressed, as molecular weight markers on Southern gels, as chromosome markers or tags to identify chromosomes, and as reagents in assays to quantitatively determined levels of GENSET expression in biological samples  
CC  
XX

SQ Sequence 163 AA:

Query Match 100.0%; Score 163; DB 6; Length 163;  
Best Local Similarity 100.0%; Pred.No.5.2e-136;  
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MGPPGFKGTGHPGLPGPKDCCGPPPGSTGPGAGGGERGAMGPGRPPGHVGP	60
DB	1	MGPPGFKGTGHPGLPGPKDCCGPPPGSTGPGAGGGERGAMGPGRPPGHVGP	60
QY	61	PGQPGAGISAVGLKGDRTGATGAGLPGQPGPPGPGPGYGMKATGPMGCGIPG	120
DB	61	PGQPGAGISAVGLKGDRTGATGAGLPGQPGPPGPGPGYGMKATGPMGCGIPG	120
QY	121	IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQOYPPMKTMGPRG	163
DB	121	IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQOYPPMKTMGPRG	163

Search completed: December 1, 2005, 08:48:14  
Job time : 188 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: December 1, 2005, 08:38:15 ; Search time 46 Seconds  
(without alignments)  
292.959 Million cell updates/sec

Title: US-09-924-340-58

Perfect score: 163  
Sequence: 1 MGPPFGKGTGHRGLPGPKG.....GAMPMEQGYPPMKTMKGFPG 163

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 572060 seqs, 82675679 residues

Word size : 100

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/1/iaa/5 COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/6 COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/H COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/PTUS COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/RE COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfileal.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	163	100.0	163	2	US-10-000-489-58 Sequence 58, Appl

## ALIGNMENTS

RESULT 1  
US-10-000-489-58  
; Sequence 58, Application US/10000489  
; Patent No. 6794363  
; GENERAL INFORMATION:  
; APPLICANT: Benjamin, Stephanie  
; APPLICANT: Tanaka, Hiroaki  
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF  
; FILE REFERENCE: 91.US6.DIV  
; CURRENT APPLICATION NUMBER: US/10/000,489  
; CURRENT FILING DATE: 2001-11-14  
; PRIOR APPLICATION NUMBER: US 09/924,340  
; PRIOR FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: PCT/IB01/01715  
; PRIOR FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: US 60/305,456  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/302,277  
; PRIOR FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/298,698  
; PRIOR FILING DATE: 2001-06-15

; PRIOR APPLICATION NUMBER: US 60/293,574  
; PRIOR FILING DATE: 2001-05-25  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: JPatent  
; SEQ ID NO 58  
; LENGTH: 163  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-000-489-58

Query Match 100.0%; Score 163; DB 2; Length 163;  
Best local Similarity 100.0%; Pred. No. 5.2e-138;  
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MGPPFGKGTGHRGLPGPKDCGKRPGRSTGRPGAEGEFGAMGPGQRPGRGHVGP	60
DB	1	MGPPFGKGTGHRGLPGPKDCGKRPGRSTGRPGAEGEFGAMGPGQRPGRGHVGP	60
QY	61	PGQPGAGISAVGLKGDGRGATGERGLAGLPQGPQPGPQGPQPGYGMKATGPMQGGI	120
DB	61	PGQPGAGISAVGLKGDGRGATGERGLAGLPQGPQPGPQGPQPGYGMKATGPMQGGI	120
QY	121	IPGPPGPMQGPQKAGHCNPSDCFGAMPMEQGYPPMKTMKGFPG	163
DB	121	IPGPPGPMQGPQKAGHCNPSDCFGAMPMEQGYPPMKTMKGFPG	163

Search completed: December 1, 2005, 08:49:50  
Job time : 47 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OW protein - protein search, using sw model

Run on: December 1, 2005, 08:48:25 ; Search time 165 Seconds  
(without alignments)  
412.765 Million cell updates/sec

Title: US-09-924-340-58  
Sequence: 163  
1 MGPPGFKGTGHPGLPGPKG.....GAMPMEQYPMKTMKPGF 163

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1867569 seqs, 417829326 residues

Word size : 100

Total number of hits satisfying chosen parameters: 8

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

Published Applications AA Main:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/us07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/us08\_PUBCOMB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/us09\_PUBCOMB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/us10A\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/us10B\_PUBCOMB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/us11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	163	100.0	163	3	US-09-924-340-58
2	163	100.0	163	3	US-09-924-340-58
3	163	100.0	163	3	US-09-924-340-58
4	163	100.0	163	3	US-09-924-340-58
5	163	100.0	163	3	US-09-924-340-58
6	163	100.0	163	3	US-09-924-340-58
7	163	100.0	163	3	US-09-924-340-58
8	163	100.0	163	3	US-09-924-340-58

#### ALIGNMENTS

RESULT 1  
US-09-924-340-58  
Sequence 58, Application US/09992600A  
Publication No. US20030027161A1  
GENERAL INFORMATION:  
APPLICANT: Benjamin, Stephanie  
APPLICANT: Tanaka, Hiroaki  
TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF  
FILE REFERENCE: 91.US4.DIV  
CURRENT APPLICATION NUMBER: US/09/992,600A  
CURRENT FILING DATE: 2001-11-13  
PRIOR APPLICATION NUMBER: US 09/924,340  
PRIOR FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: PCT/IB01/01715

PRIOR FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: US 60/305,456  
PRIOR FILING DATE: 2001-07-13  
PRIOR APPLICATION NUMBER: US 60/302,277  
PRIOR FILING DATE: 2001-06-29  
PRIOR APPLICATION NUMBER: US 60/298,698  
PRIOR FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: US 60/293,574  
NUMBER OF SEQ ID NOS: 114  
SOFTWARE: Jpatent  
SEQ ID NO: 58  
LENGTH: 163  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-924-340-58

Query Match 100.0%; Score 163; DB 3; Length 163;  
Best Local Similarity 100.0%; Pred. No. 1.2e-136;  
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPPGFKGTGHPGLPGPKDCKPSTGRPGAEGEPGAMPQGPQGPVGPVGP 60  
DB 1 MGPPGFKGTGHPGLPGPKDCKPSTGRPGAEGEPGAMPQGPQGPVGPVGP 60  
QY 61 PGQPGAGISAVGLKGDGRGATGERGLAGLPGQGPQGPQGPQGPVGMKATGPMGQGSIFG 120  
DB 61 PGQPGAGISAVGLKGDGRGATGERGLAGLPGQGPQGPQGPQGPVGMKATGPMGQGSIFG 120  
QY 121 IPQPGPMQGPQGAHCNPSDCFGAMPMEQYPMKTMKPGF 163  
DB 121 IPQPGPMQGPQGAHCNPSDCFGAMPMEQYPMKTMKPGF 163

RESULT 2  
US-09-924-340-58  
Sequence 58, Application US/09924340  
Publication No. US20030027248A1  
GENERAL INFORMATION:  
APPLICANT: Benjamin, Stephanie  
APPLICANT: Tanaka, Hiroaki  
TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF  
FILE REFERENCE: 91.US2.REG  
CURRENT APPLICATION NUMBER: US/09/924,340  
CURRENT FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: US 60/305,456  
PRIOR FILING DATE: 2001-07-13  
PRIOR APPLICATION NUMBER: US 60/302,277  
PRIOR FILING DATE: 2001-06-29  
PRIOR APPLICATION NUMBER: US 60/298,698  
PRIOR FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: US 60/293,574  
NUMBER OF SEQ ID NOS: 112  
SOFTWARE: Jpatent  
SEQ ID NO: 58  
LENGTH: 163  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-924-340-58

Query Match 100.0%; Score 163; DB 3; Length 163;  
Best Local Similarity 100.0%; Pred. No. 1.2e-136;  
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MGPPGFKGTGHPGLPGPKDCKPSTGRPGAEGEPGAMPQGPQGPVGPVGP 60  
DB 1 MGPPGFKGTGHPGLPGPKDCKPSTGRPGAEGEPGAMPQGPQGPVGPVGP 60  
QY 61 PGQPGAGISAVGLKGDGRGATGERGLAGLPGQGPQGPQGPQGPVGMKATGPMGQGSIFG 120  
DB 61 PGQPGAGISAVGLKGDGRGATGERGLAGLPGQGPQGPQGPQGPVGMKATGPMGQGSIFG 120

[illegible]

### RESULT 3

US-09-992-095B-58  
; Sequence 58, Application US/09992095B  
; Publication No. US20030157485A1

```

: NUMBER OF SEQ ID NOS: 112
: SOFTWARE: JPlatent
: SEQ ID NO: 58
: LENGTH: 163
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-992-095B-58

```

RESULT 4  
US-09-99

Sequence 58, Application No. 09/999,9570  
Publication No. US20030170628A1  
GENERAL INFORMATION:  
APPLICANT: Benjamin, Stephanie  
APPLICANT: Tanaka, Hiroaki  
TITLE OF INVENTION: HUMAN CDNA'S AND PROTEINS AND USES THEREOF  
FILE REFERENCE: G-0910US08DIV  
CURRENT APPLICATION NUMBER: US/09/999,570  
CURRENT FILING DATE: 2001-06-14  
PRIOR APPLICATION NUMBER: US 09/924,340  
PRIOR FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: PCT/IB01/01715  
PRIOR FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: US 60/305,456  
PRIOR FILING DATE: 2001-07-13  
PRIOR APPLICATION NUMBER: US 60/302,277  
PRIOR FILING DATE: 2001-06-29  
PRIOR APPLICATION NUMBER: US 60/298,698

```

; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112

```

ORGANISM: Homo sapiens  
US-09-999-570-58

Qy	1	MGPFGFKKTHPGLPGKGCGCKRGKPGSGTGRPAEGEPGAMGPOGRGPGGHVPPR	60
Db	1	MGPFGFKKTHPGLPGPKGDCKKRGPGSGTGRPAEBEPGAMGPHGRGPGGHVPPR	60
Qy	61	PGGGPGGISAVGIKKDGRATGERGLAQLPGDPGPGPGPGPGYGMGAITGSMGOGITG	120
Db	61	PGGPGPGISAVGLIKDRGATGERGLAGLPGDPGPGPGPGPGYGMKAITGSMGOGITG	120
Qy	121	IPGPPGPMGDPGKAGHCNPSCDFGAMPBEOQTPEPKTKMGPRG	163
Db	121	IPGPPGPMGDPGKAGHCNPSDCFGAMPBEOQTPEPKTKMGPRG	163

RESULT 5  
US-10-000-489-58

```

Sequence 58, Application US/10000489
Publication No. US20030092011A1
GENERAL INFORMATION:
APPLICANT: Benjamin, Stephane
APPLICANT: Tanaka, Hiroaki
TITLE OF INVENTION: HUMAN CDNA5 AND PROTEINS AND USES THEREOF
FILE REFERENCE: 91.US6.DIV
CURRENT APPLICATION NUMBER: US/10/000,489
PRIOR FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: US 09/924,340
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: PCT/IB01/01715
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: US 60/305,456
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/302,277
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/298,698
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 60/293,574
PRIOR FILING DATE: 2001-05-25
NUMBER OF SEQ ID NOS: 112
SOFTWARE: Jpatent
SEQ ID NO 58
LENGTH: 163
TYPE: PRT
ORGANISM: Homo sapiens
US-10-000-489-58

```

Query Match	100.0%	Score 163;	DB 4;	Length 163;
Best Local Similarity	100.0%	Pred. No. 1.2e-156;		
Matches 163;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY 1 MGPRGFGKKTHTPLPGCKGDCSGPRPGSGTGRPAEAEERPGAMGQSGRPGNGHPRP 60

Dd 1 MGPRGFGKKTHTPLPGCKGDCSGRPPRGSGTGRPAEAEERPGAMGQSGRPPNGHPRP 60

QY 61 PGDGPAGISAVGLKGRDGTGERGLAGLPGQDGPDPGQRPGRPGYKMGATGPMGQGIIG 120

Dd 61 PGDGPAGISAVGLKGRDGTGERGLAGLPGQDGPDPGQRPGRPGYKMGATGPMGQGIIG 120

QY 121 IPRGPRGMGDPGKAGHCNPSDCFGAMPEEOTPRPKTKMGKPRG 163

Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPF 163

RESULT 6  
US-10-000-986-58

; Sequence 58, Application US/10000986  
; Publication No. US20030096247A1  
; GENERAL INFORMATION:

; APPLICANT: Benjamin, Stephanie  
; APPLICANT: Tanaka, Hiroaki  
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF  
; FILE REFERENCE: 91.US9.DIV

; CURRENT APPLICATION NUMBER: US/10/000,986  
; CURRENT FILING DATE: 2001-11-14

; PRIOR APPLICATION NUMBER: US 09/924,340  
; PRIOR FILING DATE: 2001-08-06

; PRIOR APPLICATION NUMBER: PCT/IB01/01715  
; PRIOR FILING DATE: 2001-08-06

; PRIOR APPLICATION NUMBER: US 60/305,456  
; PRIOR FILING DATE: 2001-07-13

; PRIOR APPLICATION NUMBER: US 60/302,277  
; PRIOR FILING DATE: 2001-06-29

; PRIOR APPLICATION NUMBER: US 60/298,698  
; PRIOR FILING DATE: 2001-06-15

; PRIOR APPLICATION NUMBER: US 60/293,574  
; PRIOR FILING DATE: 2001-05-25

; NUMBER OF SEQ ID NOS: 112

; SOFTWARE: JPatent

; SEQ ID NO 58

; LENGTH: 163

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-000-986-58

Query Match 100.0%; Score 163; DB 4; Length 163;  
Best Local Similarity 100.0%; Pred. No. 1.2e-136;  
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGPPGKGTGHTGHLGPPKDDCGKPPPGSTGRPGAEGEPGANGPGORPPGHVGP 60

Db 1 MGPPGKGTGHTGHLGPPKDDCGKPPPGSTGRPGAEGEPGANGPGORPPGHVGP 60

Qy 61 PCQPGAGISAVGLKQDRGATGSRGLAGLPGQPGPPGPGPGPGYGMGATGPMGQGGIPG 120

Db 61 PCQPGAGISAVGLKQDRGATGSRGLAGLPGQPGPPGPGPGPGYGMGATGPMGQGGIPG 120

Qy 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPF 163

Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPF 163

RESULT 7

US-10-154-678-58

; Sequence 58, Application US/10154678  
; Publication No. US20030162186A1  
; GENERAL INFORMATION:

; APPLICANT: Benjamin, Stephanie  
; APPLICANT: Tanaka, Hiroaki

; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF  
; FILE REFERENCE: 182.US1.REG

; CURRENT APPLICATION NUMBER: US/10/154,678  
; CURRENT FILING DATE: 2002-10-15

; PRIOR APPLICATION NUMBER: US 09/924,340  
; PRIOR FILING DATE: 2001-08-06

; PRIOR APPLICATION NUMBER: US 60/305,456  
; PRIOR FILING DATE: 2001-07-13

; PRIOR APPLICATION NUMBER: US 60/302,277  
; PRIOR FILING DATE: 2001-06-29

; PRIOR APPLICATION NUMBER: US 60/298,698  
; PRIOR FILING DATE: 2001-06-15

; PRIOR APPLICATION NUMBER: US 60/293,574  
; PRIOR FILING DATE: 2001-05-25

; NUMBER OF SEQ ID NOS: 112

; SOFTWARE: JPatent

; SEQ ID NO 58

; LENGTH: 163

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-154-678-58

Query Match 100.0%; Score 163; DB 4; Length 163;  
Best Local Similarity 100.0%; Pred. No. 1.2e-136;  
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGPPGKGTGHTGHLGPPKDDCGKPPPGSTGRPGAEGEPGANGPGORPPGHVGP 60

Db 1 MGPPGKGTGHTGHLGPPKDDCGKPPPGSTGRPGAEGEPGANGPGORPPGHVGP 60

Qy 61 PCQPGAGISAVGLKQDRGATGSRGLAGLPGQPGPPGPGPGPGYGMGATGPMGQGGIPG 120

Db 61 PCQPGAGISAVGLKQDRGATGSRGLAGLPGQPGPPGPGPGPGYGMGATGPMGQGGIPG 120

Qy 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPF 163

Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPF 163

RESULT 8

US-10-838-854-58

; Sequence 58, Application US/10838854  
; Publication No. US20050026182A1  
; GENERAL INFORMATION:

; APPLICANT: Benjamin, Stephanie  
; APPLICANT: Tanaka, Hiroaki

; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF  
; FILE REFERENCE: 91.US6.DIV

; CURRENT APPLICATION NUMBER: US/10/838,854  
; CURRENT FILING DATE: 2004-05-03

; PRIOR APPLICATION NUMBER: US/10/000,489  
; PRIOR FILING DATE: 2001-11-14

; PRIOR APPLICATION NUMBER: US 09/924,340  
; PRIOR FILING DATE: 2001-08-06

; PRIOR APPLICATION NUMBER: PCT/IB01/01715  
; PRIOR FILING DATE: 2001-08-06

; PRIOR APPLICATION NUMBER: US 60/305,456  
; PRIOR FILING DATE: 2001-07-13

; PRIOR APPLICATION NUMBER: US 60/302,277  
; PRIOR FILING DATE: 2001-06-29

; PRIOR APPLICATION NUMBER: US 60/298,698  
; PRIOR FILING DATE: 2001-06-15

; PRIOR APPLICATION NUMBER: US 60/293,574  
; PRIOR FILING DATE: 2001-05-25

; NUMBER OF SEQ ID NOS: 112

; SOFTWARE: JPatent

; SEQ ID NO 58

; LENGTH: 163

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-838-854-58

Query Match 100.0%; Score 163; DB 5; Length 163;  
Best Local Similarity 100.0%; Pred. No. 1.2e-136;  
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGPPGKGTGHTGHLGPPKDDCGKPPPGSTGRPGAEGEPGANGPGORPPGHVGP 60

Db 1 MGPPGKGTGHTGHLGPPKDDCGKPPPGSTGRPGAEGEPGANGPGORPPGHVGP 60

Qy 61 PCQPGAGISAVGLKQDRGATGSRGLAGLPGQPGPPGPGPGPGYGMGATGPMGQGGIPG 120

Db 61 PCQPGAGISAVGLKQDRGATGSRGLAGLPGQPGPPGPGPGPGYGMGATGPMGQGGIPG 120

Qy 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPF 163

Db 121 IPGPPGPMGQPGKAGHCNPSDCFGAMPMEQYPPMKTMGPF 163

Thu Dec 1 10:27:26 2005

us-09-924-340-58.rapbm

Page 4

Search completed: December 1, 2005, 09:02:32  
Job time : 165 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: December 1, 2005, 08:49:06 ; Search time 11 Seconds  
(without alignments)  
70.954 Million cell updates/sec

Title: US-09-924-340-58

Perfect score: 163  
Sequence: 1 MGPFGFKGTGHPGLPGPKG.....GAMPMEQGYPPMKTKMPFG 163

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 26661 seqs, 4788334 residues

Word size : 100

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : Published Applications\_AA\_New:\*

- 1: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubppaa/PCT\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubppaa/US11\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	ID	Description
-----				

No matches found

Search completed: December 1, 2005, 09:02:49  
Job time : 11 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: December 1, 2005, 08:38:06 ; Search time 37 Seconds  
(without alignments)  
423.874 Million cell updates/sec

Title: US-09-924-340-58

Perfect score: 163  
Sequence: 1 MGPPGFKGTGHPGLPGRKG.....GAMPMEQYPPMKTKMGPRG 163

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 100

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : PIR 80:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query Match	ID	Description
---------------	----------------	----	-------------

No matches found

Search completed: December 1, 2005, 08:48:58  
Job time : 38 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 1, 2005, 08:28:14 / Search time 236 Seconds  
(without alignments)  
487.293 Million cell updates/sec

Title: US-09-924-340-58

Perfect score: 163  
Sequence: 1 MGPPGFGKGTGHGGLPDPKG.....GAMPMEQGYPPMKTMKGPPG 163

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 2166443 seqs, 705528306 residues

Word size : 100

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : Uniprot\_05.80:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
-----					

No matches found

Search completed: December 1, 2005, 08:45:00  
Job time : 237 secs

**THIS PAGE BLANK (USPTO)**